Trust and Adaptive Rationality

Towards a New Paradigm in Trust Research

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Sozialwissenschaften der Universität Mannheim

vorgelegt von

Stephan Rompf

Mannheim, 5. Juni 2012

Erstgutachter: Prof. Dr. Hartmut Esser

Zweitgutachter: Prof. Dr. Clemens Kroneberg

Table of Contents

Index of Tables and Figures

1. Introduction

Trust is ubiquitous. As we move through our social world, numerous encounters with other people present an opportunity for us to realize and achieve the things we want in life. The success of some of these encounters depends only on our own effort, and whether or not we can attain our goals is our sole responsibility. But in many other cases, we must rely on others, and on their good-faith attempts to do what we ask. We need to let go and give in to the risks that come with interaction, because we simply cannot control the outcome. Others may not do what we would prefer them to, and, in thus acting, they may hinder the realization of our aims, or even harm us. At times, we are conscious of these risks. Things can go wrong, yet we feel assured and secure. We nevertheless decide to take the plunge into the unknown. In some cases others disappoint us, and only then do we realize that we left ourselves vulnerable to the actions and decisions of others. This insight might come to us as a shock or surprise, and this shock brings home to our consciousness that we put ourselves into a position of vulnerability. Yet it never occurred to us to think of the risks involved in our interactions when we let go in the first place.

Both type of outcome demonstrates what common sense tells us about trust—it is sometimes very difficult, and sometimes very easy for us to trust others, and in any case it is risky. A natural question to ask is the following: how is trust warranted in the first place? There has been a recent upsurge in theoretical and empirical studies exploring the role of trust in social processes. Fueled by remarkable findings on its economic impact, the increase in research activity has sparked numerous attempts to advance our theoretical understanding of the concept of trust and its underlying mechanisms (even motivating the launch of the *Journal of Trust Research* as the first discipline-specific journal in 2011). Contributions originate from more traditional research fields, ranging from psychology and social psychology to sociology, political science, economics, law, anthropology, biology, computer science, and neuroscience. Being an interdisciplinary endeavor *par excellence*, the accumulated contributions and literature are vast. The following section serves as a short introduction of the topic to the reader. It will highlight some major insights and point to the open questions in trust research.

1.1. Achievements and Enduring Questions in Trust Research

To structure the impressive amount of knowledge at hand, it is useful to classify trust research into three categories which constitute the predominant levels of analysis: (1) microanalyses, studying the interactive generation, maintenance, and disruption of trust at the individual level, (2) mesoanalyses, investigating the effects of trust in social environments at an aggregated level; for example in dyadic partnerships, in teams, in social networks, and in organizations, and (3) macroanalyses, exploring the impact of trust on the functioning of social systems and

society at large. The rise of trust as a "hot topic" in research reflects accumulating evidence of the substantial benefits that emerge on the micro, meso, and macro levels when trust is in place.

With respect to *macrolevel* social systems, such as political or economic systems, trust is regarded as an indispensable ingredient in their smooth functioning, and in the successful circulation of the underlying symbolic media of exchange (Misztal 1996). Trust in the reliability, effectiveness, and legitimacy of money, law, and other cultural symbols warrants their constant reproduction in everyday interactions, and their aggregation into stable social structures. In essence, these modern social institutions would disappear if trust were absent (Lewis & Weigert 1985a, b). Concerning its influence on the political system and on democratic institutions, researchers have repeatedly pointed to the significance of trust as a resource that integrates and protects the underlying institutions. For example, Putnam (1993) argues that trust was a critical factor in the historic development of democratic regimes, with long-lasting effects reaching as far as present-day civic engagement.[1] Likewise, Sztompka (1996) suggests that a lack of trust was a main barrier to the successful transformation of postcommunist societies into democratic market societies, maintaining that a vital "culture of trust" is a precondition for the functioning of democratic institutions. Higher levels of trust have been associated with more efficient judicial systems, higher-quality government bureaucracies, lower corruption, and greater financial development (La Porta et al. 1997, Guiso et al. 2004). The presence or absence of trust in society can have a macroeconomic impact. Empirically, several influential studies have shown that country-level trust, along with GDP and GDP growth, are positively correlated (Knack & Keefer 1997, Zak & Knack 2001). What is more, country-specific trust predicts bilateral trade volumes and crossnational investment decisions (Guiso et al. 2004, 2009). It is no wonder that trust is regarded as an efficient mechanism governing transactions (Arrow 1974, Bromiley & Cummings 1995), a sort of "ever-ready lubricant that permits voluntary participation in production and exchange" (Dasgupta 1988: 49).

In short, trust and other forms of social capital are regarded just as important as physical capital in facilitating the creation of large-scale business organizations necessary for economic growth and the functioning of markets (Fukuyama 1995). The above studies also suggest that trust is a vital factor for the emergence and reproduction of democratic institutional arrangements, and has a critical impact on a society's political environment, its stability, economic growth, and macroeconomic outcomes.

[1] This hypothesis has been empirically scrutinized by Guiso et al. (2008), who show that historical differences between north and south Italy in the build-up of trust and social capital have translated into sizeable present-day differences in voter turnout, number of non-profit organizations, and per capita income.

Focusing on the *mesolevel*, organizational researchers have documented a substantial body of evidence revealing the stimulating effects of trust on team building and team performance, worker productivity and organizational commitment (e.g. Jones & George 1998, Dirks & Ferrin 2001, Kramer & Cook 2004). In addition, trust is related to diminished costs of interorganizational negotiation and transaction, resulting in increased revenue and turnover (Williamson 1993, Uzzi 1997, Zaheer & McEvily 1998). Regarding dyadic relationships, such as close partnerships (Rempel et al. 1985), consumer-seller relationships (Ganesan 1994, Bauer et al. 2006), and patient-physician relationships (Anderson & Dedrick 1990, Thom & Campbell 1997), trust promotes the build-up of long-term emotional attachment, the attribution of benevolent motivations and intentions, and a reduction in uncertainty, thus securing the stability of the relationship in question (Williams 2001). Being "essential for stable relationships" (Blau 1964: 64), trust is a valuable resource for individuals because, once in place, it facilitates the attainment of desired outcomes and adds to the stock of available social capital (Burt 2003). The social networks and the relations—that is, the embeddedness of actors in their social environments—constitute both a main opportunity and source of trust production (Granovetter 1985, Buskens 1998).

Considering trust on the *microlevel*, present or absent "within" the individual, it can be shown that individuals who report high levels of trust also report significantly higher levels of life satisfaction and happiness (DeNeve & Cooper 1998, Helliwell & Putnam 2004). It is no wonder that trust is generally regarded as a state worth striving for (Rempel et al. 1985, Baier 1986). It is a major factor in reducing the complexity of a contingent social life and stabilizing expectations in interactions (Luhmann 1979, 1988), and is sometimes said to be necessary even as a ground for the most routine behavior (Garfinkel 1963: 217). It enables individual cooperation, and thus promotes the further inclusion of actors into their social environment, leading to a relative advantage in comparison to low-trust types (Hardin 1993).

The question of its individual generation, maintenance, and disruption has been a prime topic of research in psychology and social psychology for over 40 years. While early research focused on the individual determinants of trust in the development of stable personality traits, and the cognitions that trust-related attributes yielded (Rotter 1967, Erikson 1989), recent research has increasingly focused on the cognitive processes involved, and on how they influence trust decisions. For instance, automatic processes may play a crucial role in the generation of trust, because salient situational features can trigger the use of trust-related heuristics and schemata (Hill & O'Hara 2006, Schul et al. 2008). Likewise, current mood influences judgments of trustworthiness (Forgas & East 2008), and humans often experience automatic emotional responses when recognizing faces and judging others' trustworthiness (Winston et al. 2002, Eckel & Wilson 2003, Singer et al. 2004, Todorov et al. 2009). In one very recent development, neuroscience studies have helped researchers to understand the neural processes

involved in the generation of trust, showing that trusting behavior is triggered by the activation of specific areas in the brain (Adolphs 2002, Krueger et al. 2007), and can be substantially modulated by neuropeptides such as the hormone oxytocin (Kosfeld et al. 2005, Zak 2005, Baumgartner et al. 2008). Generally, researchers continue to add detail to the picture of the mechanisms that generate trust on the individual level, utilizing recent advancements in social psychology and neuroscience to improve theoretical models of trust.

At the same time, the development of experimental tools for performing microlevel measurements of trust (e.g. the "trust game," Camerer & Weigelt 1988, Dasgupta 1988, Kreps 1990, and the "investment game," Berg et al. 1995) has enabled researchers to scrutinize the impact of different social institutions on the generation of trust. For instance, communication, which has long been recognized as a booster of cooperation (Isaac & Walker 1988, Sally 1995), clearly helps to promote trust (Bicchieri 2002, Charness & Dufwenberg 2006). There is consistent evidence that formal institutions, such as contracts and agreements, tend to "crowd out" intrinsically motivated trusting behavior (Malhotra & Murnighan 2002, Bohnet & Baytelman 2007, Ben-Ner & Putterman 2009), especially if they are related to punishment opportunities, or otherwise costly. Although first and third party punishment opportunities are effective in the generation of efficient outcomes (Fehr & Gächter 2000a), they prevent the build-up of mutual trust, and cooperation fails to extend to later stages of the game if these institutions cease to exist. On the other hand, a "favorable" social history and a corresponding positive reputation clearly foster the build-up of trust (Bohnet & Huck 2004, Bohnet et al. 2005). Likewise, the creation of a shared group identity (Eckel & Grossman 2005, Brewer 2008) and a decrease in social distance (Buchan et al. 2006) result in increased levels of mutual trust. Despite the progress that experimental research presents with respect to the influence of institutional arrangements on trust, these results also demonstrate the fragility of trust, proving that even minor changes in social and institutional environments may have dramatic changes on the levels of trust generated. Accordingly, a main conclusion that can be drawn from the experimental evidence is that context is critical to understanding trust (Hardin 2003, Ostrom 2003).

However, the accumulation of empirical evidence about individual decisions in social dilemmas such as the trust game has not been accompanied by equivalent progress in the development of integrative theoretical frameworks, theories, and models that would combine knowledge across disciplines (Bigley & Pearce 1998, Ostrom 2003). Despite the fact that trust research is flourishing, and many inspiring results have been uncovered, it has become almost a truism that a universally accepted scholarly definition of trust does not exist, just as no general paradigm of trust research has emerged (Lewis & Weigert 1985b, Mayer et al. 1995, Rousseau et al. 1998, Kramer 1999, Hardin 2002).

The universality and the complexity inherent in the concept can certainly be regarded as the main problem of the research (Kassebaum 2004). Since even in everyday language its meaning is multifaceted and diverse, the subsequent academic definition of and operationalization of trust is severely hampered. The concept of trust is used in a variety of distinct ways, which sometimes appear to be incompatible. For instance, definitions differ by the level of analysis, and vary with the causal role that trust is assumed to play (cause, effect, or interaction). They change with the specific context that is being analyzed, and collide when trust is viewed as static or dynamic, or conceived of as either a unidimensional or multidimensional phenomenon. In addition, trust may be confused with other concepts, antecedents, and outcomes, such as risk, other-regarding preferences, and cooperation (Mayer et al. 1995, McKnight & Chervany 2000, 2001).

The conceptual diversity of the literature on trust is mirrored in the many attempts that have been made to organize the vast interdisciplinary research (Lewicki & Bunker 1995b, a, Bigley & Pearce 1998, Rousseau et al. 1998, Kramer 1999). For example, Sitkin and Roth (1993) collect work on trust into four basic categories: (1) trust as an individual attribute, (2) trust as a behavior, (3) trust as a situational feature, and (4) trust as an institutional arrangement. Bigley and Pierce (1998) advocate a "problem-centered" approach, distinguishing between research accounts that focus on (1) interactions among unfamiliar actors, (2) interactions among familiar actors in ongoing relationships, and (3) the organization of economic transactions in general. While these approaches cut across disciplinary borders, trust research is often regarded as segregated into several traditions, which although identical on the level of observable behavior, make differential assumptions concerning the underlying mechanisms and causal elements of trust (Lewis & Weigert 1985b). Kramer (1999) contrasts the "behavioral" tradition, which principally regards trust as rational choice, to the "psychological" tradition, which attempts to understand the more complex intrapersonal states associated with trust, including a merging of expectations, affect, and dispositions. Lewicki and Bunker (1995a) differentiate a purely "psychological" tradition, which focuses on individual personality differences, from the "institutional" approach taken by economists and sociologists, and from the "social-psychological" approach, which focuses on the interpersonal transactions between individuals that generate or disrupt trust.

Conceptual dissent can even arise within the different paradigms of trust research. For instance, psychological accounts of trust usually focus on either affective *or* cognitive processes (Kassebaum 2004: 8). In economics, those accounts of trust that support a strong self-interest hypothesis (Gambetta 1988a, Coleman 1990) have been challenged by models of social preferences and a "wide" rational choice approach (see Fehr & Schmidt 2006). The result is a multitude of possible solutions for the rational explanation of trust-related phenomena. It is not surprising that a number of typologies which postulate different varieties and "types" of

trust have emerged; often limited to specific domains and research paradigms. For example, specific *versus* generalized trust (Rotter 1971, 1980), cognition-based *versus* affect-based trust (McAllister 1995), calculus, knowledge, and identification based trust (Lewicki & Bunker 1995b), and dispositional, history, category, role, and rule based trust (Kramer 1999), to name a few. In sum, "social science research on trust has produced a good deal of conceptual confusion regarding the meaning of trust and its place in social life" (Lewis & Weigert 1985b: 975), while the development of integrative theoretical frameworks has remained elusive. Historically, trust definitions have become "homonymous," preventing theoretical formulations and empirical results from becoming comparable and accumulating (McKnight & Chervany 1996).

In an attempt to reconcile this conceptual diversity, Kramer (1999: 574) argues that the presence of diverging notions of trust does not necessarily reflect insurmountable differences between incompatible models (i.e. that trust is *either* calculative *or* affective *or* role-based, etc.). Instead, a suitable theoretical framework must admit the influence of social and situational factors on the impact of instrumental and noninstrumental factors, and also articulate *how* these factors exert their influence on the decision-making process. From this perspective, future conceptualizations of trust need to integrate microlevel psychological phenomena with mesolevel group dynamics and macrolevel institutional arrangements. The interplay of individual, situational, and structural parameters and the impact of context in the development of trust have become a prime concern for research. Contributions that explicitly relate the generation of trust to internal dispositions and mental states, *as well as* to external cues and the socially structured and socially constructed environment (e.g. McKnight et al. 1998, Kramer 2006, Nooteboom 2007, Schul et al. 2008) have shifted the scholarly focus from the question "What is trust?" to the more preferable question of "Which trust, and when?"

1.2. Aim and Structure of this Work

This question is the starting point for the present study. While trust research appears to be fragmented and theoretically unintegrated, I want to show that this state of affairs has its roots in the neglect of several fundamental ingredients to trust which have not been sufficiently incorporated into the current theoretical frameworks. In emphasizing and approaching these fundamentals, the goal of the book is to equip trust research with a broad and general perspective on the phenomenon in which the conflicting perspectives and diverging types that have been previously developed can be smoothly integrated and reductively explained under a common umbrella. In short, I argue that current trust research has not sufficiently taken care of individual level *adaptive rationality*; it has furthermore failed to explicate the role of interpretation and the *subjective definition of the situation* in shaping a flexible adjustment of information processing states to the current needs of the social situation. I propose that an inte-

gration of existing trust research can be achieved along the dimension of adaptive rationality. However, this necessitates going beyond the descriptive work of creating "yet another" typology and merely sorting what has been already known. The final destination is causal explanation.

This ambition springs from the scientific approach I advocate here, which has been commonly indicated by the label of "methodological individualism" (Popper 1945, Elster 1982) and "analytical sociology" (Hedström & Swedberg 1996, 1998). In the framework of this approach, sociological explanations of collective phenomena are qualified by their focus on the, often unintended, consequences of individual actions which are restricted by structural, normative and cultural constraints and opportunities that are imposed by the social system in which the collective phenomenon emerges (Coleman 1990, Esser 1993b, 1999b). The explanative scheme, that is, the logic and structure of an analytical sociological explanation, requires that three steps be made explicit in order to understand and reductively explain a collective phenomenon: (1) a macro-to-micro transition, defining how the environment into which actors are embedded influences and restricts individual action, (2) a micro theory of action, specifying the principles by which individual actions and decisions are reached, and (3) a micro-to-macro transition, defining how a set of individual actions combine to produce a collective outcome. The combination of these three steps provides the core of any analytical nomological explanation of a collective phenomenon (Hempel & Oppenheim 1948). The present work sets out to show how the collective phenomenon of a dyadic trust relation can be analytically explained. It focuses on the phenomenon of interpersonal trust.

To this end, chapter 2 introduces the reader to the concept of trust, as defined in current trust research. The methodological device and guiding scheme to structure the review is a distinction between the *objective structure* and *subjective experience* of trust. This differentiation on the level of conceptual, empirical and theoretical description needs to be constantly observed when thinking about trust, because—as will become apparent by the end of chapter 2—an insufficient distinction between the two levels often leads to fuzzy definitions which miss precision and definitional power. Put sharply, I argue that the conversion from structure to experience (the transition from macro to micro) presents a *missing link* in trust research. Authors focus and combine different aspects of objective structure and subjective experience when defining the concept. The crucial ingredients towards linking objective structure and subjective experience—interpretation and the subjective definition of the situation—are often taken for granted, or dealt with only implicitly. By focusing research on the process of interpretation, this work seeks to contribute to the advancement of a situated cognition perspective of trust. The second chapter can be understood as an invitation to think about the necessary macro-micro link in our explanations of trust, and as an appeal for a focus on the cognitive processes involved in doing so on the level of the individual actor.

Chapter 3 continues with presenting current perspectives about the origins and explanations of trust. I review the different approaches to explaining trust in the psychological, sociological and economic disciplines. In doing so, the focus is not on emphasizing conflicts, differences and incompatibility. The basic principle in developing a broad interdisciplinary perspective is to look for the commonality, mutuality and similarity in existing research; to carve out the underlying theoretical and conceptual grounds on which a unifying theoretical framework for trust research can emerge. The chapter ends with a discussion and presentation of the main theoretical concern of the present work: the complex relation between trust and rationality. Simply put, I argue that the neglect of the dimension of rationality in the trust concept is a main barrier to the theoretical integration of existing research. While the economic paradigm assumes the capability of actors to engage in a rational, instrumental maximization of utility, sociological and psychological approaches often emphasize that trust can be nonrational and blind. Actors apply the relevant knowledge or follow cultural and normative patterns automatically, based on taken-for-granted expectations and structural assurance. Some portray trust as being based on simple heuristic processes, substituting the ideal of rational choice with a "logic of appropriateness," in which the adaptive use of rules, roles, and routines helps to establish a shortcut to trust. The discussion of the relation between trust and rationality suggests that we have to turn to other theoretical paradigms which incorporate the individual actor's degree of rationality as a fundamental ingredient. Consequentially, we will have to think about the mechanics of adaptive rationality, link it to interpretation and choice, and simultaneously maintain a clear and formally tractable model.

This task is picked up in chapter 4, which is wholly devoted to analyzing the different "routes to trust" that can be imagined when adopting the adaptive rationality perspective. It pinpoints how trustors can reach the behavioral outcome of a trusting act on different cognitive routes, linking them to different information processing states of the cognitive system. The dual-processing paradigm of social cognition will be an important resource in developing and advancing this perspective of trust. Ultimately, adopting this perspective of trust and adaptive rationality helps to resolve the enduring tensions between rational and automatic, cognitive and affective, conditional and unconditional types of trust. At the same time, it will become apparent that, in order to understand and explain adaptive rationality, we have to think about our micro theory of action, which defines and dictates how we can establish both the macro-micro and the micro-micro transition in our logic of explanation. With respect to the dual-processing paradigm, a number of factors limit its utility as an explanative vehicle in a deductive nomological explanation of trust. Most importantly, it does neither provide a causal link between interpretation and choice, nor spell out a formally precise model which can be used to guide the theoretical and empirical analysis of the trust phenomenon. This is a hindrance if the context-sensitive adjustment of rationality is assumed to be a crucial factor in determining the types of trust and the resulting subjective experiences.

I continue by introducing the "Model of Frame Selection" (MFS), a general sociological theory of action that incorporates both the aspects of a social definition of the situation and the idea of human adaptive rationality at the same time (see Esser 1990, 1991, 2001, Kroneberg 2006a, Esser 2009, 2010, Kroneberg 2011a). The MFS explicates how adaptive rationality can be formally grasped and linked to a causal explanation of action. In using the MFS to explain both conditional and unconditional types of trust, I develop a broad perspective and an integrative approach to the phenomenon. As stated before, adaptive rationality must be regarded as a key dimension of the trust concept. Moreover, the degree of rationality involved in interpretation and choice can dynamically change, it is not fixed. Thus, automatic *or* rational processes can prevail during the definition of the situation and the choice of action. The broad perspective of trust helps to integrate contradictive accounts of trust. It suggests that trust researchers have historically focused on different aspects of cognitive activity (interpretation *or* choice), assuming different processing modes (rational *or* automatic) when theorizing about a particular solution of a trust problem. The chapter closes with the development of a theoretical model that describes the *mode-selection thresholds* governing the adjustment of rationality during interpretation and choice in a trust problem.

While chapter 4 is exclusively related to developing and advancing an individual-level micro theory of action which can be used as a nomological core for an analytical explanation of trust, chapter 5 directs the reader´s attention towards the micro-macro transition which completes the last step in the logic of explanation. I scrutinize the emergence of a dyadic trust relation from an interactive and dynamic perspective in which the interpretations of the parties temporary converge into a shared *social* definition of the situation. This natural leads to adopting a perspective of reflexive structuration that draws heavily from a symbolic-interactionist interpretation of the trust phenomenon. I argue that trust is a dynamic and mutual achievement of the actors involved; it depends on *active* relational communication, identity signaling and impression management. These individual accomplishments serve as a basis for interpretation and choice because they provide the situational cues that govern information processing and trigger the activation of trust-related knowledge. Chapter 5 delineates how the trust relation, as a social system, is actively constituted by the actors involved. Far from being a passive achievement, interpretation and the subjective definition of the situation are normally reached in symbolic interaction with others, and rely on a dynamic process of communication. This also implies that the context of the trust relation cannot be static, but is highly dynamic, and endogenously shaped by the actors involved. In analyzing this last step of aggregation, the fifth chapter completes the logic of explanation of the trust phenomenon.

Overall, the theoretical part of the thesis aims at demonstrating and substantiating that a broad and integrative perspective of trust can be developed under the headnote of adaptive rationality. The modeling of the underlying general processes joins individual and social, situational

and structural factors. It allows for a causal reductive explanation of trust and an integration of the existing, but conflicting, perspectives within trust research. The explication of two missing links is necessary to establish such a framework: (1) a focus and explication of individual adaptive rationality and (2) a clarification of the role of interpretation and the definition of the situation in the trust development process. In combination, I argue, these ingredients allow for a solid explanation of different types of trust and critically advance our understanding of the phenomenon.

In chapter 6, the perspective of trust and adaptive rationality will be put to an empirical test. This test is has a twofold aim: for one, it is designed to gauge the adequacy of an adaptive rationality perspective in trust research from a general standpoint. The framing perspective of trust, as it will be developed in this work, is novel in that it merges psychological ideas of flexible information processing, "situated cognition," and a contingent use of different trusting strategies in trust problems with sociological ideas of a cultural definition of the situation and adaptive rationality. But in going beyond previous research, it specifies the causal mechanisms behind these concepts as well. The reported experiment joins them in the spotlight of empirical scrutiny. The general course of action is to operationalize and manipulate those variables which define and influence adaptive rationality. In turn, this provides a causal test of the hypotheses addressing the explanation of conditional and unconditional trust. The experiment uses the setting of an "investment game" (Berg et al. 1995). It extends this set-up with manipulations of (1) the presented context and (2) the monetary incentives, that is, "what is at stake" for the trustor. Both factors are predicted to influence the amount of rationality involved in the choice of a trusting act. Moreover, the model predicts that the internalization of trust-related knowledge shapes how these experimental factors influence trust. Overall, it points to the interactive nature of the determinants of adaptive rationality. These interactive effects and their predicted signs are one of the main concerns of the present empirical study. While the hypotheses are certainly more prone to falsification, they also portray the high empirical content of the presented theory, and go beyond the statement of simple main effects.

1.3. Summary of Empirical Results

The development of the empirical test requires deriving hypotheses which predict the sign of the expected statistical effects (see chapter 6.3 for details). The MFS carries a very important general message: a number of determinants, such as cognitive motivation, chronic and temporary accessibility, or the presence of relevant situational cues, influence the degree of rationality involved in interpretation and choice. However, their effect can only be analyzed jointly, and, depending on the concrete value of another determinant, one factor may or may not become important in a particular trust problem. As will be demonstrated, the model of trust and adaptive rationality, in conjunction with a set of experiment-specific auxiliary hypotheses, can

be used to derive a closed set of admissible *interaction patterns* that are predicted to emerge in a statistical model when analyzing the experimental data. These patterns are a specific feature and consequence of the adaptive rationality perspective. Their prediction attests to the high informational value and empirical content of the theoretical model, and they carry a number of important propositions about the trust development process.

Principally, the model suggests that trustors need not always be rational and rely on a controlled processing of information when defining a trust problem and making a choice. Instead, if favorable conditions prevail, interpretation and choice may unfold rather automatically, guided by a "logic of appropriateness," in which a number of shortcuts to trust can be used. Trust then may be based on feelings, on heuristic shortcuts, or on an unconditional execution of social norms, roles, routines, and all sorts of trust-related scripts. The MFS suggests that automatic action selection is most likely if actors have strongly internalized a corresponding script, and if the social situation points to its appropriateness and applicability, that is, if knowledge and situation "match" with each other. In this circumstance, actors may follow a script and maintain an unreflected routine even if the potential costs of an error and a wrong decision are very high.

One of the most noteworthy empirical results is the finding of such an interaction between incentives, the internalization of trust-related knowledge, and the context in solving a trust problem. I argue that previous studies of monetary incentive and stake size effects have not controlled for the factor of chronic frame and script accessibility, which is an important mediator of cognitive motivation in the model of adaptive rationality. In line with the predictions generated here, I can show that the degree of rationality involved in trust highly depends on the internalization of trust-related scripts (i.e. the norm of reciprocity). The degree of norm internalization counterbalances the negative effects of monetary incentives on trust, which, if left unchecked, would push trustors towards a rational and controlled consideration of the trust problem. In other words, trustors who have strongly internalized a trust-related script may be "less rational" and chose unconditional trusting strategies more often than those who have no access to a corresponding script. Similarly, I find that the context influences trust and the degree of rationality involved: if cues suggest the validity of trust-related frames and scripts, then trustors readily use this information during the choice of a trusting act. In short, a cooperative context can be sufficient in suppressing incentive effects as well. This also means that incentives and context do interact on a basic level, a finding that is important insofar as previous experimental studies have usually assumed their independence.

In addition, the analysis of recorded decision times reveals an overall coherent picture. That is, the predicted interaction patterns which I derive from the model match over the domain of two different dependent variables. This result is most noteworthy in itself; it lends strong support to the adequacy of the adaptive rationality perspective on trust, and the MFS in general.

For example, a high degree of norm internalization leads to a decrease in decision times. This indicates a shift to more automatic processing of information, and the prevalence of unconditional trusting strategies, for those trustors who have internalized trust-related norms. Moreover, if much is at stake and the situation involves potentially high costs of error, this increases decision times, indicating a shift to more conditional trusting strategies. But importantly, the second effect is mediated by chronic script accessibility, and it disappears with high norm internalization. Moreover, decision times decrease in a cooperative context, indicating a shift towards unconditional trusting strategies. I find that this effect depends on, and is mediated by, the accessibility of trust-related knowledge. At the same time, I show that decision times in the context of the trust problem are also strongly dependent on context free processing preferences, as measured and controlled for in the form of "faith in intuition" and "need for cognition" (Epstein et al. 1996). This is a remarkable finding for trust research, as it helps to clarify the looming tension between intuitive and rational approaches to trust, which are an ever-present facet of theorizing. The current data support a perspective in which individual differences in processing preferences shape the mode-selection threshold over and above the influence of situational and social factors and determine the resulting trusting strategies and the resulting type of trust.

Taking things together, the empirical test shows that two different behavioral indicators of conditional and unconditional trusting strategies (that is, observed levels of trust *and* corresponding decision times) can be explained with the help of one general theoretical model. The discovery of matching patterns and their similarity over the domain of two different dependent variables strongly prompts to the adequacy and validity of the adaptive rationality perspective of trust. The chapter ends with a discussion of the study´s limitations, potential caveats and unresolved questions.

Chapter 7 presents a synthesis of the proposed theoretical and empirical framework that was developed in this thesis. It delineates the broad perspective on trust, summarizing the main conclusions and propositions, the benefits and pending problems in adopting this integrative perspective, and it re-connects the current work to the broader research agendas of social science. I consider the perspective of trust and adaptive rationality to be a most useful guide for our study of the trust phenomenon because it directs our attention towards the cognitive processes involved in suspension and the "leap of faith," but it tackles them in a reductive logic by pointing to the underlying general cognitive processes. It also reminds us that a successful explanation must go beyond the creation of typologies and descriptive work, and instead master the difficult requirement of causal modeling, to which the present work seeks to provide a starting ground. The thesis ends with a discussion of open questions and avenues for future trust research.

2. The Concept of Trust

" ... trust is a term with many meanings" (Williamson 1993: 453).

To begin a scientific conceptualization of trust, we can ask, "What does the term *to trust* really mean?" In answering this question, we reveal the term's diverse and equivocal use in vernacular language. We may "trust" other people with respect to their future actions, or "trust" organizations with respect to the promised quality of their products. We "trust" a doctor when we see her to cure us, as well as with respect to her abilities and intentions to heal us. When driving in traffic, we "trust" others to abide by the rules, just as we do. Some people "trust" the government, while others only "trust" in god or in themselves. Obviously, in each example, the term "to trust" refers to a different situation and to a different object, and connotes a different meaning. In fact, it is impossible to uncover a consistent and universal notion of trust based on the everyday usage of the term (McKnight & Chervany 1996). Some researchers argue that the analysis of ordinary language is a futile tool in trust research, not only because it cannot promote one meaning of "trust" above all other candidates, but also because the use of the term—and even its very existence—varies greatly between different languages (Hardin 2002: 58). On the other hand, if scientific definitions are too distant from their everyday counterparts, they run the danger of missing important dimensions of the concept under scrutiny, and such definitions should therefore be informed by common-sense understanding (Kelley 1992). This is especially true for the concept of trust, which has a breadth of meaning in everyday life.

Unsurprisingly, trust researchers commonly relate scientific conceptualizations of trust to their everyday counterparts, for example by comparing dictionary and scientific definitions (Barber 1983, McKnight & Chervany 1996), by explicitly assessing lay theories of trust and individual-qualitative experiences (Henslin 1968, Gabarro 1978, Kramer 1996, Weber & Carter 2003), and by analyzing the meaning of the term in everyday language (Baier 1986, Lahno 2002). Most importantly, these studies suggest that trust must be conceived of as a multidimensional concept which (1) develops only under certain structural conditions and (2) merges different cognitive, affective, and behavioral dimensions into a unitary social experience (Lewis & Weigert 1985b, Rempel et al. 1985, McAllister 1995). We will therefore organize our introductory study of the concept along these two fundamental ingredients of trust: the objective structure and the subjective experience of trust. As it turns out, this course of action is most helpful in delineating the core problems and contradictions that mark the current state of trust research.

To begin with, a reasonable level of consensus exists on the structural prerequisites which form the set of conditions necessary for trust to arise (Rousseau et al. 1998: 395). We will re-

fer to these as the *objective structure* of trust. Most importantly, the situation must involve a risk which stems from the trusting parties' uncertainty about the preferences of the trusted party. Secondly, the situation must be marked by social interdependence, meaning that the interests of one party cannot be achieved without reliance on another party. As a corollary, the trusting party will have to transfer control over certain resources or events to the trusted party (Coleman 1990) and thereby become objectively vulnerable (Heimer 2001). Lastly, trust has to be future-oriented, in the sense that the outcome of trust cannot readily be observed, but will be determined at a more or less specified point of time in the future (Luhmann 1988).

In contrast, the *subjective experience* of trust—that is, the internal mental state associated with trust—seems to be one prime reason for the "confusing potpourri" (Shapiro 1987: 625) of trust definitions in the literature. Although there appears to be a substantial consensus on defining trust primarily as a mental state, there is much less agreement when it comes to the precise definition thereof. The propositions that have been offered by researchers are contradictory (Bigley & Pearce 1998). In a crossdisciplinary review of 60 articles on trust, McKnight and Chervany (1996) find that 50% of scientific definitions include cognitive elements, such as expectations, beliefs, and intentions, while 37% include affective elements, such as feelings of confidence and security. Almost all relate trust to some form of action. About 60% of all definitions locate trust on more than one dimension. What is more, some authors conceive of trust of as a state that is explicitly *not* perceived until it is broken, as something "non-cognitive" (Becker 1996), related to aspects of automatic decision making (Hill & O'Hara 2006, Schul et al. 2008) and preconscious processes which shape perception (Zucker 1986, Luhmann 2000, Lahno 2002). In a nutshell, while trust has clear structural antecedents, its subjective experience is a source of disagreement among trust researchers, because different experiential phenomena come into focus. As a result, the acknowledged phenomenology of trust and its resulting definition differ remarkably between research traditions and disciplines.

This poses a challenge to the development of a comprehensive theory. If the phenomenological foundation of trust is variable, then it is no surprise that its subsequent conceptualization and definition remain contradictory. Different authors focus on different phenomenological aspects of the same explanandum. As a consequence, trust definitions have historically become too narrow and "homonymous," preventing theoretical formulations and empirical results from becoming comparable and accumulating (McKnight & Chervany 1996). Although everyday language cannot provide a unitary precise definition, it does help trust researchers to sharpen their conceptualizations, because it suggests how the phenomenology of trust should be conceived of. As it is, analysis of ordinary language indicates that the subjective experience of trust is in fact multifaceted (for example, consider the use of the term in idioms such as "I trusted you blindly!" versus "Trust, but verify!"). Ultimately, in moving towards a broad theoretical framework, we will have to absorb this wide phenomenological

foundation of trust, and to explain when and why the subjective experience of trust differs so vastly between situations, as well as which factors (internal and external) promote, constrain, and shape the subjective experience of trust.

The following chapter presents a comprehensive review of the core concepts of trust research. Moving from objective structure to subjective experience, the "ingredients" of trust are systematically explored, thereby also revealing the challenging diversity in the trust definitions present in the literature. Even though the amount of research reviewed may initially appear confusing and contradictory, this exercise is nonetheless most rewarding, as it enables the general lines of conflict, the stumbling blocks, and bones of contention in trust research to be carved out. The chapter includes a delineation of the conceptual boundaries between trust and related concepts, such as confidence, system trust, distrust, and the like. As it is, the concept of trust is often used only in conjunction with a host of related concepts. Some of these are antecedents to trust; others only seem to be related, and are frequently confused with trust. The discussion of the conceptual boundaries completes the introduction of the trust concept, and rounds up the terminology that will be used throughout the book. The chapter closes with a discussion of the relation between objective structure and subjective experience. It is argued that a prime reason for the diversity and contradictions among trust definitions is rooted in the fact that trust researchers rarely pay attention to the process of interpretation—the subjective definition of the situation—and how it relates to the objective structure of the trust problem. The central question that this chapter opens up is, "how does the objective structure of the trust problem translate into the subjective experience of trust?"

2.1. Objective Structure

2.1.1. Constituents of Interpersonal Trust Relations

Trust is a phenomenon that we usually ascribe to other persons or to ourselves, but not to "things" or inanimate objects. Although we have not yet dealt with the question of the subjective experience of trust, it is clear that we can always find *ex post* reasons which have motivated a trusting act, even if that trust was misplaced. In other words, individual and purposeful *actors* are the primary *subjects of trust*. This raises the question of whether collective actors can be a subject of trust as well. The "US government" might trust the "North Korean Regime" with respect to nuclear policies, for example. We could even extend the concept to include collective "trust systems" (Coleman 1990) and the variations of collective action they enable. As Coleman notes, "the analysis of these phenomena requires going beyond the two- or three-actor systems, but it can be done through the use of these components as building blocks" (ibid. 188). The analysis of collective actors, collective trust systems and collective action is, however, beyond the scope of this book. The present work seeks to conceptualize and understand trust as an outcome of individual framing and decision-making processes. If

trust is conceived of as a mental phenomenon, then the individual microlevel represents its sole level of emergence. Hence, the concept will be limited to individual actors and their dyadic relations in the following. As a subject of trust, these actors will be called *trustors*.

Trust generally denotes a special relation between two actors, but it can additionally refer to a relation between a trustor and other groups, organizations, or more abstract social institutions. Other actors, groups, organizations, institutions, and the like are examples of primary *objects of trust*. Many researchers have developed ideal-type trust classifications to refine these basic distinctions, for example by varying the degrees of social distance and generalization of the objects of trust (Bigley & Pearce 1998, Sztompka 1999). But given the broad experiential basis that trust can assume, any attempt to identify, classify, and validate various ideal-type objects of trust must remain incomplete and arbitrary (Möllering 2001). The case of *interpersonal* trust between two actors can nevertheless be regarded as a prototypical case. Dyadic trust relations form the micro social building blocks of larger systems of trust—this makes their comprehension a premise for understanding a wide range of social phenomena, including social integration at large. Focusing on these building blocks, the concept of trust will in the following be limited to a relation between two individual actors. Trust in groups and collective actors, in more abstract institutions (such as expert knowledge systems), and trust in the political or economic system as a whole will not be subject to detailed analysis. These types of trust will be delimited from the concept of interpersonal trust under the rubric of "system trust." They are relevant insofar as they can become the basis for certain types of interpersonal trust that use system trust as a starting point (a topic which will be more fully explored in chapter 3.2).

The special relation between two actors to which trust refers will be called a *trust relation*. Its emergence is the prime explanandum of this work. Trust relations are always a three-part relation in the form of "A trusts B with respect to X" (Hardin 2002: 9). In a trust relation, the second actor B—the object of trust—will be called the *trustee*. The item X is called the *content of the trust relation*. Trustor, trustee, and the content of the trust relation are the main constituents of interpersonal trust relations. Note that, for the moment, we restrict the trust relation to a one-way relation. The notion of trust is often connected to the idea of a situation of mutual trust and iterated exchange. These are aspects of "social embeddedness" and of corresponding two-way trust relations in a social environment, which will be introduced later. Interpersonal trust, even when it is conceptualized as a psychological state of the trustor, is always relational and social in the sense that the trust relation necessarily extends to the dyad and "transcends" the boundaries of the individual (Lewis & Weigert 1985b).

2.1.2. The Basic Trust Problem

The emergence of a trust relation is contingent upon certain structural prerequisites which confront the trustor with a particular decision-making problem, when he has to decide *whether or not to trust*. In the basic *trust problem*, a trustor faces two different sets of actions:

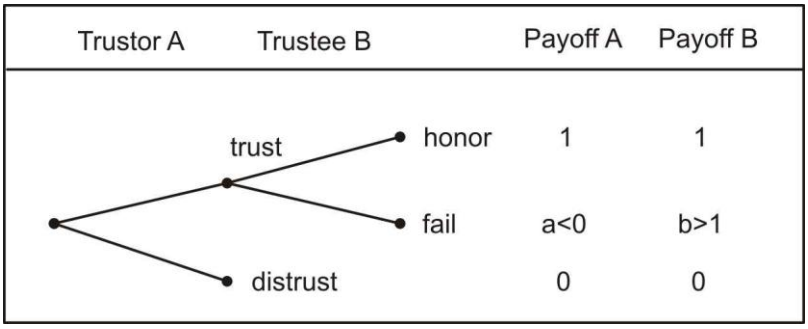
(1) Actions from the first set make him vulnerable with respect to the actions of the trustee. The trustor transfers control over resources or events to the trustee, and the trustee's future actions determine whether the trustor will experience a loss or a gain. While making his decision, the trustor cannot foretell with certainty how the trustee will decide, and he cannot rely on sanctions or any other form of external enforcement to induce the desired outcome.

(2) In contrast, actions from the second set allow the trustor to eliminate potential damage, with certainty and from the beginning. In this case, the trustor does not put himself into a position where the trustee can determine the loss or gain. By refraining from a transfer of resources or control, the trustor is not taking a risk, and he prevents getting into a vulnerable position; he can maintain the *status quo* with certainty.

If a trustor chooses actions from the first set, we say that "*A trusts B*," and the observable action will be called a *trusting act*. The choice of a trusting act constitutes the trust relation between the actors. If A chooses actions from the second set, then "A *distrusts* B," and the observable action is distrust.

It is now possible to spell out more precisely the content X of a trust relation: If "A trusts B with respect to X," then all classes of actions which (1) do not harm A, and (2) serve to realize the prospective gain and utility increase for A, belong to the content of the trust relation X. If B chooses his actions accordingly, he *is trustworthy*, he *fulfills* A's trust, or he *honors* A's trust. Note that a trustworthy response often leaves some "latitude of judgment" to the trustee. That is, the content X may be very specific, and may demand a unique course of action, or it may be more general, defining the desired outcome state, but not the precise actions necessary to achieve it. Mutual gains can be achieved through trust and trustworthy response, but there are also incentives for the trustee to choose the untrustworthy option and to fail A's trust (Messick & Kramer 2001: 91). Principally, a trustworthy response requires some form of effort (time, energy, or other resources) and thus has a cost to the trustee which he can save by simply not fulfilling the content of the trust relation. Deutsch (1958) emphasizes that trustworthiness implies that the trustee will fulfill the content of the trust relation, even if violating trust is more immediately advantageous. The interests of the trustor are violated if the trustee disregards the content of the trust relation. In that case, "B *fails* A's trust" and violates the trust relation. Coleman emphasizes that a breach of trust must put the trustor in a worse situation than if he had not trusted (Coleman 1990: 98f.). The following picture summarizes the basic trust problem (figure 1):

Figure 1: The basic trust problem



In this case, the *status quo* payoffs are zero for both actors, the successful establishment of a trust relation yields payoffs (1|1), respectively. A failure of trust puts the trustor into a worse position as compared to *status quo*, while the trustee experiences some gain that puts him in a better position than a trustworthy response. To conclude, the trustor has to decide whether to trust the trustee with respect to X. The content of the trust relation comprises all actions which improve A's utility with respect to the status quo and "realize" the content of the trust relation. These prospective gains present a basic motivation to engage in the trust relation. However, the transfer of control over resources or events involves the risk of incurring a loss if the trustee disregards the content of the trust relation and fails A's trust. These prospective losses present a basic risk. While mutual gains can be achieved from trust and trustworthy response, there is also an incentive for the trustee to fail the trust. Taken together, these structural conditions constitute the basic trust problem.

2.1.3. Trust and Action

As we have seen, if "A trusts B with respect to X," then the trustor chooses an action which makes him vulnerable with respect to the actions of the trustee. On the level of overt behavior, we can observe the choice of a trusting act, manifested as a transfer of control over resources or events to the trustee. In other words, the behavioral content of trust is a risky course of action through which trust is demonstrated (Lewis & Weigert 1985b). The choice of a trusting act can be interpreted as a "risky investment" (Luhmann 2000: 27), because the trustor must transfer control over certain resources or events to the trustee, and at the same time is incapable of determining the final outcome (the potential gain or loss) with certainty. So far, we have used the notion of a "choice" that the trustor "decides" on without further explication. But the issue is not trivial, and warrants closer inspection: how are trust and choice related? Is "trusting" equal to "making a choice"?

In the "behavioral approach" to trust (Kramer 1999), researchers define trust solely in terms of cooperative choices in an interpersonal context (e.g. Deutsch 1958, 1960, Loomis 1959). Trust is defined as a behavioral outcome based on sufficiently positive expectations which allow the

trustor to choose a risky course of action (Gambetta 1988a, Coleman 1990). Essentially, trust is regarded as a rational choice among actions. For instance, Gambetta suggests that trusting someone means that "the probability that he will perform an action that is beneficial to us ... is high enough for us to consider *engaging in some form of cooperation* with him" (1988a: 217, emphasis added). One advantage of this approach is that it opens up the toolbox of rational choice theory, which then can be applied to the study of trust decisions (Messick & Kramer 2001: 91). Defining trust as choice behavior also warrants that it can be examined with the help of experimental devices (Fehr 2009: 238).

Most authors maintain, however, that trust fundamentally differs from the choice of a trusting act. The trusting act merely displays the observable behavioral outcome of trust. In essence, "the fundamental difference between trust and trusting behaviors is between a 'willingness' to assume risk and actually 'assuming' that risk" (Mayer et al. 1995: 724). Pointing to the question of choice, Hardin notes, "Trust is in the category of knowledge, acting on trust is in the category of action ... I do not, in an immediate instance, choose to trust, I do not take a risk. Only actions are chosen" (2001: 11ff.). In this line, authors often conceive of trust as a "mental phenomenon" (Lahno 2002: 37) and a "psychological state" (Kramer 1999: 571). It is intrinsically tied to the subjective categories of knowledge and affect. When trusting someone, the trustor—in some way—accepts the vulnerability inherent in a trust problem. The choice of a trusting act merely displays an observable outcome.

Do we "choose" to trust, then, or not? This far from trivial question will be subject to analysis throughout the book. As we will see, trust can be understood as a two-step process, which begins with interpretation—that is, with the subjective definition of the situation—and leads to (and prepares for) a subsequent choice of action. The choice of the trusting act can approximate a rational decision-making process, but it need not. As such, "choosing to trust" can be an appropriate empirical description, but it can also be misleading if automatic processes prevail. To keep the terminology simple, we will use the notion of choice ("if A chooses to trust B," "if A trusts B," "the choice of a trusting act" etc.) to denote the fact that a trustor has opted for a transfer of control and is in a vulnerable position—ignoring, for the moment, the way in which this outcome has been internally achieved by the trustor.

A closely related question is whether trust and cooperation are the same (as suggested, for example, by Gambetta's definition, cited above). Although observable cooperative choices and trust are intimately related—cooperation can be a manifestation of trust—it is problematic and confusing to simply equate the two concepts (Good 1988, Mayer et al. 1995, Hardin 2001). Cooperation may occur for many reasons, even when there is no risk taken, no potential loss at stake, or no choice available. The most extreme case imaginable may be a situation in which cooperation is enforced by deterrence and the threat of punishment, against the will of the actors involved. While this, to an alien observer, might look like some form of coopera-

tion, it is clearly not an outcome of trust. Therefore, whenever we observe cooperation, we must carefully ascertain whether or not we can ascribe trust to the actors involved in the interaction. Conceptually, these are not the same. As it is, cooperation must be conceived of as one indicator, among many others, of the latent construct of trust (McKnight & Chervany 1996: 32). In fact, researchers have recently devised means of experimentally separating trust from cooperation, showing that an upward spiral of benign attributions and increased cooperation is involved in the mutual build-up of trust (Yamagishi et al. 2005, Ferrin et al. 2008).

2.1.4. Social Uncertainty

Assume a state of perfect information. If a trustor knew the trustee's preferences, corresponding motivations, and intentions with certainty, he could predict whether the trustee would honor or fail his trust. In this deterministic setting, there would be no "need" for trust, although it would nonetheless be possible to observe trusting acts. Since the trustor can determine the outcome and knows all preferences and incentives, his actions can more adequately be described as *reliance* (Nooteboom 2002). Imagine, on the other hand, that no information at all was available; then the trustor could just as well roll a dice to make the decision in a trust problem. In this setting, there would be no "opportunity" for trust, although again we could observe trusting acts. In this case, it is more adequate to speak of *hope* (Lewis & Weigert 1985b, Luhmann 2000: 28). Trust is limited to instances where specific knowledge structures play a crucial role in the solution of a trust problem (Endress 2001: 175). In short, the concept of trust addresses a state of knowledge that is neither perfect, nor completely ignorant (Simmel 1992: 392).

In a trust problem, information is asymmetric and imperfect. While certain characteristics of the trustee (his preferences, motivation, and intentions) are hidden to the trustor, they are perfectly well-known to the trustee. At the same time, the trustor usually has at least some information that he can use in a given trust problem—but this information is typically imperfect. In the case of imperfect information, the likelihood of an event (for example, "trustworthy response to trust") can be assessed as, at best, some probability. We will henceforth refer to the subjective assessment of the probability of an event as an *expectation*. Depending on the "precision" of his expectations, a trustor subjectively faces either a situation of *risk* or of *ambiguity*. The terms expectation, risk, and ambiguity will be treated as subjective categories, and they describe the internal representations of the objective uncertainty involved in the trust problem (see chapter 2.2.2).[1]

[1] Note that some authors (e.g. Knight 1965) have used the term "uncertainty" to refer to a subjective state of "ambiguity" (Camerer, Weber 1992: 326). The terminology adopted here follows Camerer & Weber (1992) and minimizes the risk of confusion: Uncertainty is *objective*, while risk and ambiguity are *subjective*, as discussed in chapter 2.2.2 below.

What is the source of uncertainty for the trustor? In a trust problem, uncertainty is based on the fact that the trustor cannot control the outcome of the trusting act once he has chosen to trust, and also on his imperfect information about the trustee. Since the outcome genuinely rests upon his interaction with the trustee, the trustor faces a fundamental *social uncertainty* (Kramer 2006: 68). Social uncertainty is endogenous to interactions. It results from the contingent decisions of other actors, and becomes relevant whenever the utility of an actor is directly or indirectly influenced by the decisions of others. As we have seen, this is exactly the case in trust problems. Unlike exogenous, environmental uncertainty, an actor can to some degree influence endogenous, social uncertainty. For example, a trustor can mitigate social uncertainty by opting for the safe alternative of distrust. However, he can never avoid it when he chooses a trusting act. Social uncertainty is a constitutive element of the basic trust problem.

All in all, trust problems are characterized by social uncertainty and asymmetric, imperfect information. The trustor has limited information about the trustee's preferences, motivations, and intentions and, when confronted with the choice of a trusting act, cannot be sure of the future actions of the trustee. As noted by many researchers, the mix of social uncertainty and imperfect information, paired with the possibility of opportunistic action, is a core element of trust problems (Dasgupta 1988, Gambetta 1988a, Mayer et al. 1995, Luhmann 2000, Hardin 2001, Heimer 2001, Kramer 2006). The fact that these elements are structural prerequisites and objective "facts" does not, however, give an answer to the question of how trustors subjectively handle the uncertainties involved in a trust problem. This question is, as we will see, the crux of trust research, and it will be dealt with throughout the remainder of the book. Most trust researchers agree that trust is a special way of dealing with social uncertainty and imperfect information. In choosing to trust, the trustor—somehow—bypasses the social uncertainty inherent in the trust problem and takes a "leap beyond the expectations that reason and experience alone would warrant" (Lewis & Weigert 1985b: 970). We can conclude at this point that trust hints at the particular nature of the expectations involved, a particular way that they emerge and form, and a particular way of dealing with the uncertainty and the subjective risk or ambiguity involved in a trust problem.

2.1.5. Vulnerability

The concept of trust is almost routinely linked to the aspect of vulnerability (Hosmer 1995, Bigley & Pearce 1998). In fact, almost all research on the topic of trust rests on the idea that actors, in some way or other, become vulnerable to each other during their interaction. Many authors specify this vulnerability by referring to the objective structure of trust. From this perspective, vulnerability simply means that something must be "at stake" for the trustor. Trust always includes a transfer of control, and therefore results in the objective vulnerability of the trustor (Deutsch 1960, 1973). In the absence of vulnerability, the concept of trust is not neces-

sary, as outcomes become irrelevant to the trustor (Mishra 1996). Vice versa, vulnerability increases with the proportion of total wealth that is at stake in an interaction (Heimer 2001). In short, vulnerability originates from the interaction in a basic trust problem, and it mirrors its incentive structure.

In contrast, many authors emphasize the importance of the subjective perception of vulnerability for trust, and use the term only with reference to the trustor's subjective experience. In contrast to a structural prerequisite or mere consequence thereof, the term vulnerability then describes a qualitative element of individual subjective experience—an internal response to the incentive structure. In this perspective, the most commonly emphasized elements are favorable expectations and the willingness and intention to be vulnerable. For example, Rousseau et al. propose that "trust is a psychological state comprising the *intention to accept vulnerability* based upon positive expectations of the intentions or behavior of another" (1998: 395, emphasis added). That is to say, when facing a trust problem a trustor consciously perceives vulnerability, but nevertheless chooses the risky course of action. This "behavioral-intention subfactor" of trust (Lewicki et al. 2006), in addition to cognitive and affective elements, constitutes the heart of many trust definitions.

A number of scholars argue explicitly against such a linkage of trust and vulnerability. They emphasize that vulnerability remains outside of the trustor's awareness, even when it objectively exists. Vulnerability becomes salient to the trustor only after the trustee has failed trust; the conscious perception of vulnerability is linked to a psychological state of distrust (Becker 1996, Lahno 2001, 2002, Keren 2007, Schul et al. 2008). On the other hand, trust is linked to a state of inner security and certainty in which potential risks are not consciously experienced. This apparent discrepancy in trust definitions is one example of a problematic merging of objective-structural and subjective-experiential components in definitions of trust (see section 2.4 below). Disregarding the question of its subjective experience for the moment, we first have to establish at this point that vulnerability is always manifest in the objective structure of a trust problem, and becomes tangible by the transfer of control.

2.2. Subjective Experience

2.2.1. The Phenomenology of Trust

In a broad review of the trust literature, Kramer notes: "Most trust theorists agree that, whatever else its essential features, trust is fundamentally a psychological state" (1999: 571). It is now our task to understand how the phenomenological foundation of trust might be specified. As a psychological state, trust points to a special way of dealing with the social uncertainty and vulnerability inherent in trust problems. Although there seems to be a substantial consensus on treating trust as a mental phenomenon, scholars are less in agreement on what precisely

characterizes this state of mind (Bigley & Pearce 1998). The aspect of its qualitative subjective experience is widely debated, and scholars hold somewhat diverse views.

Many researchers emphasize that an intentional and conscious acceptance of vulnerability and the voluntary taking of risk are necessary, in order for us to be able to speak of trust (Deutsch 1958, Luhmann 1988, Mayer et al. 1995, Rousseau et al. 1998). For example, cognitive conceptualizations of trust focus on the expectations, beliefs, and intentions which a rational utility-maximizing actor uses to make a decision in a given trust problem. Trust from this perspective is a purely cognitive phenomenon, marked by a retrieval of existing knowledge, a resulting "cold" unemotional expectation, and a corresponding rational choice of action. In the worst case, the existing knowledge has to be updated, but essentially, trust remains in the category of knowledge (Hardin 2001).

Psychological studies often put a special emphasis on affective aspects of trust. For example, the emotional bonds between individuals can form a unique basis for trust, once developed (McAllister 1995), and both mood and emotions can color the subjective experience of trust, signaling the presence and quality of trust in a relationship (Jones & George 1998). Trust, it is argued, establishes commitment, generates a feeling of confidence and security, and induces attachment to the trustee (Burke & Stets 1999). Some researchers maintain that in a state of trust, we do not perceive social uncertainty and vulnerability at all. Instead, the subjective perception of risk or ambiguity is effectively suppressed and replaced by a feeling of certainty and security that lasts until trust is failed (Garfinkel 1967: 38-52, Baier 1986, Becker 1996, Jones 1996, Schul et al. 2008). For example, Baier suggests that "most of us notice a given form of trust most easily after its sudden demise or severe injury" (1986: 234). In these contributions, trust is not merely a matter of "cold" expectations, but refers to a "hot" affective state, or to preconscious processes which filter our perception (Holmes 1991, Lahno 2001).

But the role of affect and emotions in creating trust is relatively unexplored in comparison to the large number of cognitive accounts that have been proposed. What is more, trust researchers seldom pay attention to the interaction of cognitive, affective, and behavioral elements, and how they jointly determine possible forms of trust that can emerge in the course of interaction (Bigley & Pearce 1998). Moreover, it is debated whether trust can be genuinely characterized as a state (Lagerspetz 2001), or whether it is expressed "punctually" in different situations and decisions (Luhmann 2000: 34).

Overall, trust is acknowledged to be a complex, multidimensional phenomenon which must be defined in terms of interrelated processes and orientations, involving cognitive, affective, and behavioral elements (Lewis & Weigert 1985a, b, Bigley & Pearce 1998, Kramer 1999). Note that the term *cognitive* is used in a narrow sense here, to include higher mental functions such as thinking, reasoning, judgment, and the like. On the other hand, the term *affective* here re-

fers to experiences of feeling and to emotional sensation and arousal, charged with a positive or negative valence. The proposed dimensionality of the trust construct closely mirrors the classic trichotomy of attitude research, which proposes affective ("feeling"), cognitive ("knowing"), and behavioral ("acting") dimensions of attitudes (Breckler 1984, Chaiken & Stangor 1987). Unsurprisingly, trust has, by some researchers, been defined as an *attitude* towards the object of trust (e.g. Luhmann 1979: 27, Jones & George 1998, Lewicki et al. 2006).

Presumably trust and its subjective experience have a "bandwidth" (Rousseau et al. 1998: 398) and can take various forms in various relationships, or even within the same relationship. Conceptualizations range from a calculated weighing of perceived gains and losses, to an emotional response based on interpersonal attachment and identification. One reason for the widely divergent views on the subjective experience of trust is that its distinct cognitive, affective, and behavioral manifestations need not necessarily be present at one point in time (Lewis & Weigert 1985a). The influence of each dimension varies with the specific trust relation. It is easy to imagine that the affective dimension of trust will be more pronounced in close relationships, while the cognitive dimension will be more influential in trust relations with secondary groups and strangers, or in market-based exchanges. Although these dimensions are analytically distinct, the experience of trust always includes all of them: cognitive elements, such as knowledge structures and the corresponding expectations; affective elements, such as feelings, moods, and emotions; a behavioral intention to act on trust; and an implicit reference to the normatively and culturally structured social environment, which all merge into the unique experience of trust. Essentially, "trust in everyday life is a mix of feeling and rational thinking" (Lewis & Weigert 1985b: 972).

2.2.2. Expectations and Intentions

One of the earliest and most commonly emphasized elements in academic definitions of trust is that of a favorable expectation about the outcome of a trusting act (Deutsch 1958, Rotter 1967, Zand 1972, Barber 1983, Lewis & Weigert 1985b, Gambetta 1988a, Coleman 1990, Hardin 1993, Robinson 1996, Rousseau et al. 1998, Kramer 1999, Möllering 2001, Lewicki et al. 2006). In a trust problem, a trustor uses his available knowledge to form "a set of expectations, assumptions, or beliefs about the likelihood that another's future actions will be beneficial, favorable, or at least not detrimental to one's own interests" (Robinson 1996: 576). We will summarize and label these "expectations, assumptions, or beliefs" as the *expectation of trustworthiness* of the trustor. According to Gambetta (1988a), this expectation is located on a probabilistic distribution with values between complete distrust (0) and complete trust (1), with a region of indifference (0.5) in the middle. The choice of a trusting act then requires that an expectation exceeds a subjective threshold value. As Gambetta points out, "optimal" threshold values will vary as a result of individual dispositions and change with situational circumstances. Once the expectation of trustworthiness exceeds the threshold and is high

enough to engage in the choice of the trusting act, we will say that it is a *favorable* expectation.

Rotter (1967) defined interpersonal trust as a favorable expectation held by an individual that the word, promise, or oral or written statement of another individual or group can be relied upon. In addition, he proposed a distinction between *specific* and *generalized* expectations. While specific expectations are based on experiences in a situation with a unique individual, and rest on a concrete interaction history, generalized expectations are abstracted and synthesized from a plurality of past experiences in similar situations. In a given situation, individuals use both specific and generalized expectations to assign intentions and motives to interaction partners. The influence of generalized expectations increases with the novelty and unfamiliarity of the situation (Rotter 1971). Expectations provide a "lay theory" to the trustor of what he can expect in a given situation in response to his choice of a trusting act, and they represent the "good reasons" which constitute the evidence of trustworthiness. These are part of his socially learned, though always imperfect, knowledge, and they allow a judgment to be made of the trustee's trustworthiness.

The idea that trust is based on an expectation points to the important fact that the trustor, in some way or the other, has to make an inference about the trustworthiness of the trustee. In doing so, he uses available information and queries the trust-related knowledge he has stored in his memory. The main questions that cognitive accounts of trust provoke naturally is, "What are the sources of trust-related knowledge?"; "On which indicators can the inference be based?" and "How is this trust-related knowledge actually used?"

As it is, there are many sources of trust-related knowledge. Some may pertain directly to characteristics of the trustee; others reflect indirect sources, such as generalized expectations or knowledge about institutions and the like. In general, one can distinguish "macro" sources, which apply generally and impersonally, from "micro" sources, which arise in specific exchange relations and are personalized (Nooteboom 2006). A related distinction between macrosources and microsources has already been made by Deutsch (1973: 55), who differentiated between "universalistic," "generalized," and "particularistic" expectations of trustworthiness. In a similar fashion, Zucker (1986) identified three "modes" of trust production, and tied them to particular categories of trust-related knowledge: "process-based" trust is founded in repeated experience with a trustee or his reputation, "characteristics-based" trust relies on an assessment of stable characteristics of the trustee, while "institution-based" trust refers to a set of shared expectations derived from formal social structures. Likewise, Sztompka (1999) defines "axiological" trust as relating to an assessment of the trustee's predisposition to follow normative rules, and "fiduciary" trust as relating to an assessment of the trustee's inclination to meet moral obligations, whereas "instrumental" trust is based on assessments of the trustee's competence and past performance. All in all, trust researchers commonly refer to both

microsources and macrosources of knowledge when thinking about the informational basis of trust.

Conceptualizing interpersonal trust in dyadic trust relations, a most frequent focus is on the characteristics of the trustee that promote a favorable expectation of trustworthiness (Butler & Cantrell 1984, Rempel et al. 1985, Baier 1986, Sitkin & Roth 1993, Hosmer 1995, Mishra 1996, McKnight & Chervany 2000, for a review see Mayer et al. 1995). These characteristics refer to individual assessments of the trustee's personality and thus, generally speaking, belong to the formation of specific expectations. Each characteristic contributes a unique perspective from which a trustor can consider the trustee and the trust relation. As such, trustee characteristics also help us to learn more about the quality of the subjective experience of trust. They are considered, among other factors, to be "antecedents of trust" (Mayer et al. 1995: 727). From the many characteristics proposed in the literature, four major characteristics can be extracted.[2] In sum, the perceived (1) benevolence, (2) competence, (3) integrity, and (4) predictability of the trustee are assumed to notably shape perceived trustworthiness and the resulting level of interpersonal trust. Trustee characteristics can be thought of as subcategories of the higher-level construct "expectation of trustworthiness." Adopting a slightly different terminology, McKnight et al. use the term "trusting beliefs" to refer to the "secure conviction that the other party has favorable attributes, such as benevolence, integrity, competence and predictability" (2006: 30). We will use the terms "belief" and "expectation" synonymously in the following.

(1) *Benevolence* indicates an assumption concerning the trustee's motivation. The trustee is benevolent to the extent that he cares about the welfare of the trustor, and therefore respects the content of the trust relation. That is, a benevolent trustee will not choose the opportunistic act of failing trust, even if there are incentives to do so, and has the intention of acting in the interest of the trustor. In the context of dyadic trust relations, benevolence points to a personal orientation of the trustee towards the trustor; it is the perception of his goodwill, caring, and responsibility. Benevolence also points to the normative dimension of trust. It is itself a moral value (Hosmer 1995), and implies a responsibility or "fiduciary obligation" (Barber 1983) to care for the protection of the trustor.

(2) *Competence* or *ability* refers to the capability and qualification of the trustee to fulfill the content of the trust relation. In order to honor trust, the trustee must possess certain skills to successfully complete actions which realize the content of the trust relation. This may require technical competencies, interpersonal competencies, the ability to make judgments, and so forth. The domain-specificness of skills and competencies carries forward to the trust concept

[2] Following Mayer et al. (1995) and McKnight et al. (1998).

itself (Zand 1972). That is, one primary reason why A trusts B with respect to X, but not with respect to Y, is that expectations of competence and ability do not extend over all domains.

(3) *Integrity* means that the trustee adheres to a set of principles which the trustor finds acceptable. The degree to which a trustee is judged to have integrity will be influenced, for example, by the consistency of the trustee's past actions; by his honesty, truthfulness, and openness in communication, by the extent to which a trustor believes the trustee's statements of future intentions, and by how congruent his actions are with his words (Hosmer 1995). Perceived integrity is always "value-laden," because the trustor evaluates the perceived trustee's moral and behavioral principles within his own value system (McKnight & Chervany 2000). Again, this points to a normative dimension of the trust concept. Integrity is influenced by the degree of perceived value congruence between trustor and the trustee (Sitkin & Roth 1993). Integrity will be high if the trustor can identify a shared value or moral principle and assume that the trustee acts accordingly (Jones & George 1998).

(4) *Predictability*. While integrity denotes a value-laden evaluation of the trustee's personality, the aspect of predictability refers to a value-free perception of the trustee's consistency in action. Predictability means that the actions of the trustee are consistent enough over time so that a trustor can forecast to a satisfying degree what the trustee will do. Although a perfect opportunist may not be judged to have high personal integrity, his actions are predictable. With high predictability, other characteristics of the trustee, such as his benevolence or competence, do not vary over time. However, predictability alone is insufficient to make a trustor willing to take a risk and choose a trusting act: if a trustee, with a high degree of predictability, will act opportunistically, then a trustor may still withhold trust, if he is convinced that the trustee will seek his own advantage by not being trustworthy.

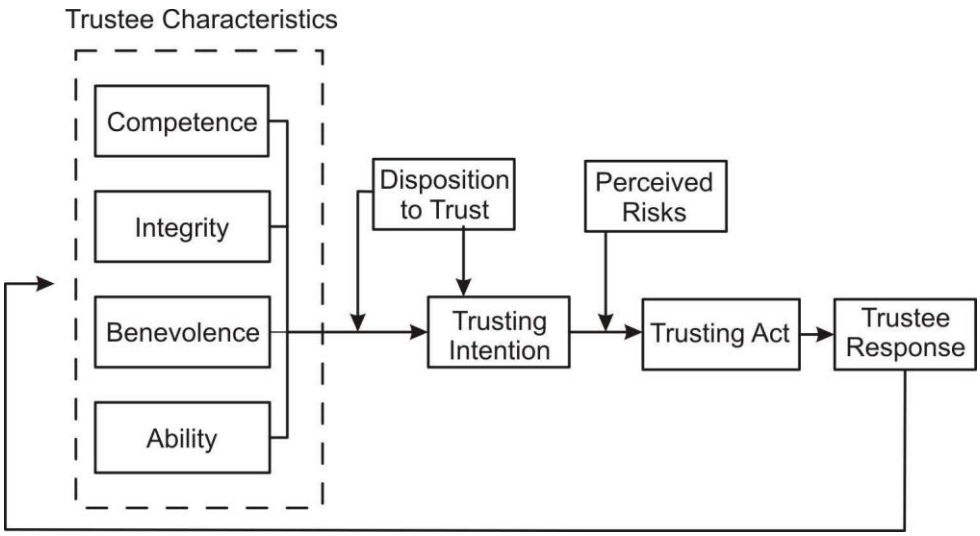
According to Mayer et al. (1995), these characteristics jointly explain a major portion of perceived trustworthiness. In fact, evidence of their absence often provides a rational basis for withholding trust (Shapiro 1987, McAllister 1995). Mayer et al. (1995) have developed a widely received model of trust development in which these trustee characteristics, in combination with a trustor's internally stable and generalized "disposition to trust,"[3] wholly determine the level of trust an actor has. Importantly, they define trust as "the *willingness* of a party *to be vulnerable* to the actions of another party *based on the expectation* that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (ibid.: 712, emphasis added). Note that this definition focuses on trust as an intentional consequence of expectation formation ("willingness to be vulnerable"), rather than on

[3] Mayer et al. (1995) use the term *disposition to trust* to refer to generalized expectations of trustworthiness. These are understood as referring to stable personality traits, which influence how much trust one has "prior to data on that particular party being available" (ibid. 715); see also chapter 3.1.3.

the act of trusting. Similarly, McKnight and Chervany define a *trusting intention* as "a secure, committed willingness to depend upon, or become vulnerable to, the other party" (2006: 30), and separate it analytically from the expectation of trustworthiness. Formally, they claim, expectations and intentions must be treated as distinct subcategories of the high-level construct of trust; expectations are regarded as the causal antecedents to trusting intentions (McKnight et al. 1998, also Ferrin et al. 2008). Note that these contributions emphasize the subjective experience of vulnerability, that is, a conscious acceptance of risk, and the intentionality inherent in the choice of a trusting act.

In the model presented by Mayer et al. (1995), the trusting intention is compared to the level of perceived risk in a given trust problem (notably, this resembles Gambetta's definition of a subjective threshold value to which actual expectations are compared). If intentions are sufficiently strong, the trustor engages in "behavioral risk taking" by choosing the trusting act (see figure 2):

Figure 2: The model of trust development, adapted from Mayer et al. (1995: 715)

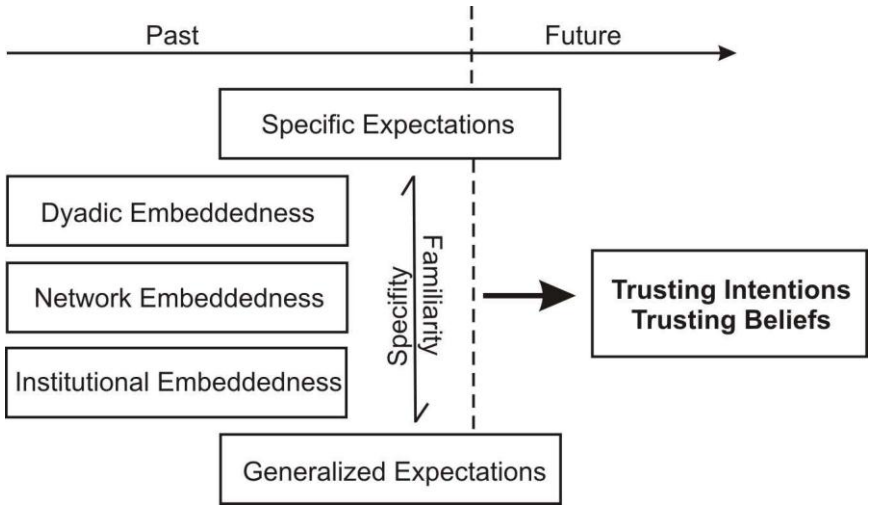


Several points of about this model are worthy of comment. First, the judgment of trustworthiness and the evaluation of risks are assumed to rely on available information. Thus the model presents a purely cognitive approach to trust, viewing it basically as the outcome of a rational inference process (Schoorman et al. 2007). Second, the model is wholly focused on unidirectional interpersonal trust. That is, it captures neither trust in more abstract institutions and social systems or the potential influence of other macro sources, nor the development of mutual trust. Most importantly, there is no explicit reference to the social environment in which the trust relation is embedded, and no reference to other categories of trust-related knowledge that may motivate trustworthiness. The model is "dynamic" through feedback from the outcome (the response of the trustee) to the input factors of perceived trustworthiness, and therefore allows for repeated interaction. However, it is "static" in the sense that, in a given trust prob-

lem, there is no reference to communication or interactive processes through which the parties involved "define" their perspectives to negotiate perceived trustworthiness (see Jones & George 1998, Bacharach & Gambetta 2001, Nooteboom 2002; this issue will be taken up in chapter 5). Fourth, even when contextual factors are assumed to change the levels of perceived risk and trustworthiness, trust essentially remains a joint function of trustee characteristics and generalized expectations. Although the proposed causal relation between trusting beliefs and trusting intention appears to be empirically justified (McKnight & Chervany 2006, Colquitt et al. 2007, Schoorman et al. 2007), the assumed cognitive basis of trust is clearly very narrow.

Considerable attention has been paid by trust researchers to identify further sources of trust-related knowledge which the trustor can access in a given trust problem (e.g. Granovetter 1985, Zucker 1986, Shapiro 1987, Buskens 1998, Jones & George 1998, McKnight et al. 1998, Ripperger 1998, Kramer 1999, Luhmann 2000, Möllering 2006a, Buskens & Raub 2008). Many theoretical accounts demonstrate a much broader cognitive basis for trust and a plethora of "good reasons" on which the trustor can base his decision. Conversely, they present a much broader motivational basis for a trustee to be trustworthy. In a nutshell, the most important micro and macrosources of trust-related knowledge derive from (1) specific expectations, including knowledge of characteristics of the particular trustee and/or the structure of the particular, unique trust problem; (2) the dyadic embeddedness of the trust relation, that is, knowledge from repeated and reciprocal interaction; (3) the network embeddedness of trustor and trustee, that is, knowledge of the social mechanisms of reputation, learning, and control; (4) institutional embeddedness and the internalization of trust-related norms, social roles, cultural practices, as well as knowledge concerning institutional safeguards such as legal contracts, regulations, and guarantees; (5) generalized expectations, individual "dispositions to trust," moral principles, values, stereotypes, and the like. These categories will be introduced and discussed in depth in chapter 3. For now, it is sufficient to note that a trustor can utilize a very broad base of informational resources that help to establish expectations in a trust problem (see figure 3):

Figure 3: Sources of trust-related knowledge



In sum, cognitive accounts define trust as the willingness and intention to be vulnerable based on sufficiently favorable expectations of trustworthiness in the face of social uncertainty (Mayer et al. 1995, McKnight et al. 1998, Rousseau et al. 1998). The trustor's stored knowledge provides the foundation and cognitive basis, from which a transition into expectations, intentions, and action can be made. The trustor uses his trust-related knowledge to form expectations, to develop intentions, and finally to decide whether the choice of a trusting act is justified or not. As Coleman argues, situations of trust can be viewed as a particular subset of those involving risk: "The elements confronting the potential trustor are nothing more or less than the considerations a rational actor applies in deciding whether to place a bet" (Coleman 1990: 99). Note that this implies a perception of social and environmental uncertainty in terms of subjective *risk*. Although the trustor faces endogenous and exogenous uncertainty, he can synthesize the available information into a precise estimate of the risk involved in the choice of a trusting act. In doing so, the trustor relies on a broad base of informational resources, such as trustee characteristics, generalized expectations, and knowledge of the social environment in which the trust relation is embedded, which enable him to take an appropriate course of action.

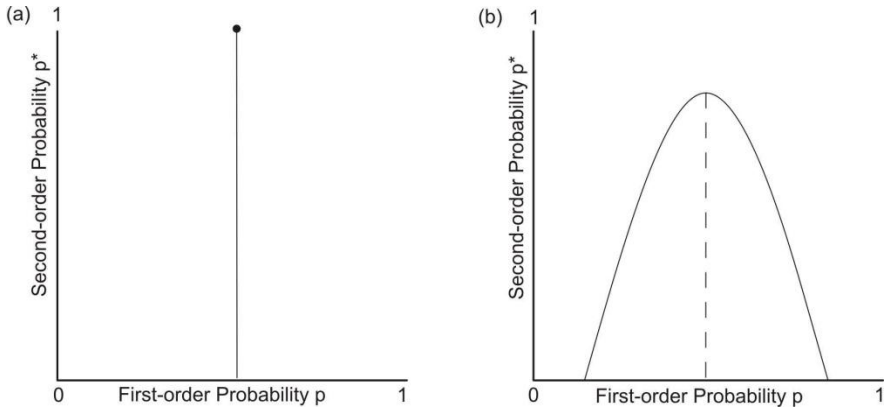
2.2.3. About Risk

At this point, it is necessary to take a closer look at the relation between expectations, uncertainty, risk, and ambiguity (see Frisch & Baron 1988, Camerer & Weber 1992). According to scholarly definitions, both risk and ambiguity refer to the subjective perception of objective uncertainty. In a situation of *risk*, a decision-maker can predict the occurrence of events with a precise probability. The risk of a situation is represented by his expectations; it can be summarized as lotteries over outcomes, which are the formal apparatus for modeling risk (Mas-Colell et al. 1995). Risk can be understood as the perceived probability of loss which originates from

the choice of an uncertain event. It includes opportunity costs in the form of foregone gains that result from disregarding other alternatives (Chiles & McMackin 1996). In a situation of risk, the utility of an actor itself becomes risky, and assumes the form of *subjective expected utility* (Savage 1954, Anscombe & Aumann 1963). Most importantly, risk describes the assessment of a precise subjective probability, and therefore is also called "unambiguous probability" (Ellsberg 1961). For example, an event e_i may or may not occur, but it does so with an exact subjective likelihood of $p(e_i)$. In short, in a situation of risk, the occurrence of an event is uncertain, but the expectation of its occurrence is unambiguous.

In contrast, *ambiguity* characterizes a state of knowledge that is not sufficient to even make a "good guess." With ambiguity, it is not possible to attach precise probabilities to events. Instead, the corresponding expectations are "ambiguous," distributed along a second-order probability distribution (Marschak 1975, Einhorn & Hogarth 1986). When ambiguity increases, the second-order probabilities become flatter, or more evenly distributed around the mean. In consequence, expectations become more ambiguous.[4] In the worst case, the second-order probabilities become equally distributed. Then, the actor has no information at all about the probability of an event; he faces a state of complete ignorance. On the other hand, if there is absolutely no ambiguity, then expectations resemble point estimates from the second-order probability distribution, with a variance of zero and a second-order probability of one: the situation is reduced to a situation of risk. Put differently, risk refers to precise expectations of uncertain events, while ambiguity refers to the imprecision of expectations. The terms risk and ambiguity refer to the subjective perception of objective, endogenous, or exogenous uncertainty (figure 4).

Figure 4: Risk (a) and ambiguity (b)



Returning to the problem of interpersonal trust, note that the preceding exposition offers a host of interpretations concerning the question: "How is trust related to uncertainty, risk, and

[4] Keeping the terminology consistent, we will *not* say: "Expectations become more uncertain"!

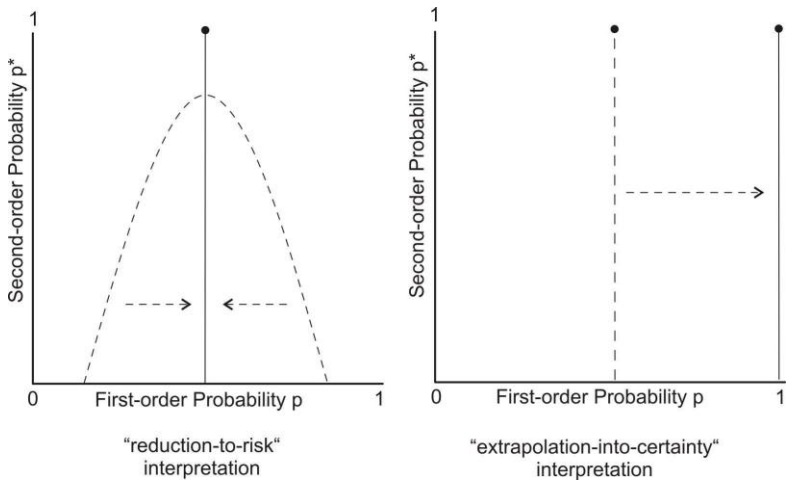
ambiguity?" To begin with, in most cognitive accounts, trust is assumed to be based on perceived risk and on the corresponding expectations in the given trust problem. Taking these authors by their words, this implies that there is no ambiguity present when a trustor decides to trust. In using his available knowledge, a trustor converts all uncertainty inherent in a trust problem into subjective risk. Most cognitive models of trust share the assumption that trust is a matter (or "a subset") of risk, and thus, one of unambiguous expectations. We can add to this argument by allowing for the presence of ambiguity at the initial stage of the trust problem. Imagine a situation in which the trustor is in a state of complete ignorance, with correspondingly flat second-order probabilities. Every piece of information that the trustor can use to interpret the situation will help him to establish more unambiguous expectations. By considering his broad base of knowledge, the second-order distributions become denser and denser, until they finally converge to point estimates—expectations of trustworthiness are then stable and unambiguous: the subsequent choice of a trusting act is possible. In essence, trust is solely related to the categories of knowledge, expectation, and risk (Coleman 1990, Hardin 2002).

However, such an interpretation fails to inform us of the purported "cognitive leap beyond the expectations that reason and experience alone would warrant" (Lewis & Weigert 1985b). As we have conjectured at the outset, trust characterizes the particular nature of the expectations involved, a particular way of their emergence and formation, and a particular way of dealing with risks and ambiguity in a trust problem. In a purely cognitive approach, there is nothing "beyond" the knowledge that promotes trust. But according to Lewis and Weigert (1985b), "to trust is to live *as if* certain rationally possible futures will not occur" (ibid. 969). Möllering (2001) introduces the idea of *suspension* to capture this "as-if" notion of trust. He claims that suspension is a process that "brackets out" ambiguity by making the available knowledge momentarily certain. This facilitates the transition into favorable or unfavorable expectations. More pointedly: trust comes into play exactly when available knowledge is *not* sufficient to stabilize otherwise ambiguous expectations.[5] Applying this argument figuratively, a trustor would use his knowledge and available information to reduce ambiguity towards stable expectations, but at some point, further reductions would become no longer possible. Now, by suspension, all remaining ambiguity is converted into manageable risk. Unfortunately, Möllering does not offer an explanation of the mechanisms behind suspension. But he makes the important point that interpretation—that is, the trustor's subjective definition of the situation—is the constitutional ground on which expectations, as an output of interpretation, are built. In order to understand suspension, he concludes, it is necessary to better understand the process of interpretation.

[5] Similarly, Lewis and Weigert conclude that " ... knowledge alone cannot be a fully adequate basis for the expectations informing social action. Trust begins where knowledge ends" (1985a: 462).

Furthermore, if suspension is part of the trust phenomenon, it can be brought to an interesting extreme. Imagine that, given a lack of further supportive information, suspension not only reduces ambiguity down into risk, but instead completely eliminates risk *or* ambiguity by replacing them with *subjective certainty*. In other words, the effect of suspension could also be an "extrapolation" of (un)ambiguous expectations into the extremes of subjective certainty. This entails the subjective experience of security and self-assurance if available information is extrapolated into the direction of a certain and favorable expectation of trustworthiness with an attached probability of one. And it entails a clear perception of distrust when the extrapolated expectations become fixed at zero. Such a perspective coincides with theoretical accounts that characterize trust as a state in which risks are *not* perceived. In a similar fashion, Luhmann asserts that trust replaces external uncertainty with internal certainty by means of "overdrawing" (2000: 30ff.) information. More pointedly: while risk implies the consideration of a set of uncertain alternatives, trust implies their elimination (Luhmann 1979: 25).[6] As Luhmann claims, trustors cannot usually access their knowledge in the form of expectations in a particular trust problem, and risk need not be part of the trustor's subjective experience at all, before the choice of a trusting act is made. From this perspective, a "cognitive leap" resembles the suppression of perceived risks and the "extrapolation" of any form of subjective uncertainty into subjective certainty (see figure 5).

Figure 5: Potential effects of suspension on cognitive expectations and trust



All in all, cognitive accounts of trust focus on the choice of a trusting act, in which the trustor, quite similarly to one undertaking a bet, uses his expectations to rationally evaluate a lottery which grants him some expected (dis)utility. Trustors are aware of the risks, weighing the prospective gains and losses in order to choose the alternative with the highest expected utili-

[6] " ... the benefit and rationale for action on the basis of trust are to be found ... in, and above all, a movement towards indifference: by introducing trust, certain possibilities of development can be excluded from consideration. Certain dangers which cannot be removed but which should not disrupt action are neutralized" (Luhmann 1979: 25).

ty. As Coleman puts it, situations of trust are a "subset" of those involving risk (Coleman 1990: 99f.). However, the process of expectation formation is controversial. Expectations may be formed by the retrieval of appropriate knowledge, but our discussion of the relation between risk and ambiguity—as well as the introduction of the idea of suspension—reveal that trust may also rely on a further element in which "our interpretations are *accepted* and our awareness of the unknown, unknowable and unresolved is *suspended*" (Möllering 2001: 414). This notion of suspension adds a noncognitive or "irrational" element to the cognitive conceptualization of trust, which allows trustors to go beyond their expectations. It can be interpreted either as suspension of ambiguity into unambiguous risk, or as an extrapolation of risk or ambiguity into subjective certainty. Both interpretations illustrate the idea that trust enables a "cognitive leap" into stable expectations during the process of interpretation, so that the trustor can make a decision just *as if* certain possible futures will not occur.

2.2.4. Morals of Trust

Many definitions of trust seem to be based, at least implicitly, upon the idea that a trustee has a moral obligation to behave in some "justified" manner. If a trustor willingly becomes vulnerable to the actions of the trustee, he does so with a favorable expectation that the trustee will respect the content of the trust relation and act accordingly. This does not only involve favorably assessing the ability and competence required for the fulfillment of the content of the trust relation: Scrutinizing trustee characteristics for their normative substance, we find that characteristics such as benevolence, goodwill, honesty, and integrity all invoke strong normative standards of how a trustee "ought" to behave. The reason for the strong normative notions in many trust definitions is simple: trust is problematic essentially because the trustor experiences social uncertainty with respect to the moral qualities of the trustee.[7] Trustworthiness necessitates a voluntarily adherence to the content of the trust relation, and it requires—at least— restraint from opportunistic action. As such, it cannot be dealt with without asking for the moral qualities of the trustee.

The notion of an assumed "benevolent," "fair," and "justified" action on the part of the trustee is a recurring theme in the trust literature. In an attempt to gauge the scope and dimensionality of trustworthiness expectations, Barber (1983: 9f.) concludes that expectations of morally correct performance and the fulfillment of "fiduciary duties," that is, direct moral responsibilities, are as important as expectations of technical competence, and adds them to the set of expectations that motivate the choice of a trusting act. Bromiley and Cummings (1996) define trust as an expectation that another individual or group will "(a) make a good faith effort to behave in

[7] As Dasgputa puts it: "The problem of trust would of course not arise if we were all hopelessly moral, always doing what we said we would do in the circumstances in which we said we would do it. This is, the problem of trust would not arise if it was common knowledge that we were all trustworthy. A minimal non-congruence between individual and moral values is necessary for the problem of trust to *be* a problem" (1988: 53).

accordance with and keep commitments, (b) be honest in negotiations preceding those commitments, and (c) not take excessive advantage of others even when the opportunity for opportunism is available" (ibid. 303). What is more, these "trusting expectations," they argue, differ from purely "calculative expectations" because it is not assumed that the trustee's fundamental motivation is the pursuit of self-interest (Bromiley & Harris 2006). Mishra (1996) expresses this idea by adding the dimension of *concern* to trustworthiness expectations. Concern goes beyond believing that another party will not be opportunistic—it means that self-interest is balanced by the trustee's interest in the welfare of the trustor. Messick and Kramer (2001) define trust "as making the decision *as if* the other person or persons will abide by ordinary ethical rules that are involved in the situation" (ibid. 91). They argue that the two most important of these rules involve telling the truth and avoiding harming others, although other rules may become relevant in a particular situation as well. Taken together, the normative dimension of trust widely informs scholarly definitions; it is a constant undercurrent in most theoretical works.

Hosmer (1995) defines trust as "the expectation by one person, group, or firm *of ethically justifiable behavior*—that is, morally correct decisions and actions based upon ethical principles of analysis—on the part of the other person, group, or firm in a joint endeavor or economic exchange" (ibid. 399, emphasis added). This definition links trust directly to the subject of normative ethics and morality. According to Hosmer, trust is always accompanied by an assumption of an acknowledged and voluntarily accepted moral duty to protect the interest of the trustor. That is, from the trustor's perspective, the choice of a trusting act constitutes a "psychological contract" (Robinson 1996),[8] that includes a moral *obligation* to respond trustworthily (Dasgupta 1988, Lahno 2002). This obligation is more than a promise to avoid harm. It amounts to an unspoken guarantee that the interests of the trustor will be included in the final outcome. As Hosmer argues, however, no generally accepted rule exists as to how a trustee can combine and balance the conflicting interests in a "fair," "justified," or "benevolent" manner. Since the objective structure of trust does not include direct external enforcement mechanisms *per se*, the trustor has to rely on the trustee's adherence to *moral principles* (Hosmer 1995). Moral principles define what is considered a "fair," "justified," and morally correct response by the trustee.[9] They can be understood as rules that underlie decision making and restrict the pursuit of individual self-interest (Vanberg 1994: 42f.). Thus, the choice of a trusting act indicates that the trustor expects a morally correct response, and a morally cor-

[8] "A psychological contract emerges when one party believes that a promise of future return has been made [content of the trust relation], a contribution has been given [transfer of control] and thus, an obligation has been created to provide future benefits [trustworthy response]" (Robinson & Rousseau 1994: 346).

[9] Hosmer presents ten moral principles from the tradition of moral philosophy, such as universal rules (Kant), distributive justice (Rawls), utilitarianism (Bentham, Mill), and personal virtues (Plato, Aristotle).

rect response requires an action informed by a moral principle which restrains the pursuit of self-interest.

Similarly, Jones and George (1998) propose that the psychological construct of trust is not only experienced through expectations, but through *values*, such as loyalty, helpfulness, fairness, benevolence, reliability, honesty, responsibility, integrity, and competence. Values are general standards that are intrinsically desirable ends for actors. They are incorporated into a larger value system by prioritizing them in terms of relative importance as guiding principles (Rokeach 1968, 1973). Values are intricately related to moral principles: while values describe an ideal desirable end state, moral principles present the rules of how such an ideal end state is to be achieved. Consequently, most of the content of value systems is *de facto* rule-based (e.g. "do not cheat," "keep your promises," "do not harm others"). Since they are socially shared, rule-based value systems also constitute *social norms* regarding acceptable behavior (Messick & Kramer 2001). Furnishing criteria that an actor can use to evaluate events and actions, value systems guide behavior and the interpretation of experience. They allow a judgment of which behaviors and types of events, situations, or people are desirable or undesirable.

According to Jones and George (1998), this means that the perception and evaluation of trustee characteristics and the formation of trustworthiness expectations is dependent upon the trustor's value system. A trustor whose value system emphasizes fairness and helpfulness, for example, will strive to achieve fairness and helpfulness in trust relations with others, and will evaluate others primarily through these values. A favorable evaluation of the trustee depends on the degree of perceived value congruence between the trustor's and the trustee's value systems (ibid. 535f.). However, this does *not* mean that trust is possible only when trustor and trustee have completely similar values. More specifically, as long as there are no obvious signs for value *in*congruence, a trustor discards the belief that the trustee may have values that are different from his own. As Jones and George put it: "The actor ... simply *suspends* the belief that the other is not trustworthy and behaves *as if* the other has similar values and can be trusted" (1998: 535, emphasis added).[10] In short, the motive and intention that the trustor attributes to the trustee when he in fact chooses to trust is acknowledged as morally justified and valuable (Lahno 2001). The choice of the trusting act expresses the recognition of a shared value, results in the suspension of any further doubts, and maintains and reinforces the shared value at the same time (Barber 1983, Jones & George 1998). On the other hand, an actual perception of value incongruence fosters unfavorable expectations of trustworthiness and can quickly lead to distrust (Sitkin & Roth 1993).

[10] Although they introduce the idea of *suspension* to their conception of trust, Jones & George (1998) do not elaborate on this process any further.

The common characterizing feature of these normative accounts of trust is that the motivation for trustworthiness is assumed to be rooted in the trustee's compliance to moral principles, values, or other social norms (Hardin 2003). That is, a trustor trusts because he thinks that the trustee has a moral commitment to be trustworthy, upholds values that are considered as justified, or follows a social norm such as the norm of reciprocity. Hardin correctly notes that such accounts are primarily concerned with normatively motivated trustworthiness. He admits that such motivations will be relevant in many trust problems, but remarks that they do not present the only "good reasons" for a trustor. Shapiro (1987) criticizes the normative perspective as proclaiming an "oversocialized" version of the regular trustor and trustee; a view that draws too heavily on the generalized morality of individuals. Hence, it is important to keep in mind that moral dispositions of the trustee are just *one* reason that may motivate trustworthiness, and *one* source for favorable expectations among many others.

The normative undertone that sounds in many trust definitions suggests that trust is something more than a simple, "cold" expectation. Trust implies a promise of a future reciprocal transaction in the form of a trustworthy response. This is because the trustor, by his choice of a trusting act, signals that he recognizes the trustee's character as worthy of his trust. The implicit demand of a morally correct response creates an obligation for the trustee to prove his trustworthiness, and to honor the risk that the trustor takes by relying on the trustee's character. Of course, this necessitates that the moral values which create such obligation are in fact shared among the two parties. In short, trustworthiness often is not only "expected," it is normatively "required" and regarded as something "good" in itself (Hardin 2001: 21). A trustor who chooses to trust sees himself as *entitled* to the right of a fair and justified treatment. A violation of trust disappoints this demand, which may be motivated by moral principles, ethical values, or social norms.

Importantly, the normative dimension of trust also draws our attention to the social systems in which the trust relation is embedded, to the social norms prevailing within them, and to the values promoted by the larger superstructures of society. It points to a strong linkage of individual, institutional, and cultural elements in the phenomenon of trust. That is, trust cannot be fully understood and studied exclusively on either a purely individual or a collective level, because it thoroughly permeates both (Lewis & Weigert 1985b). Generally speaking, where external enforcement mechanisms are not feasible, internal mechanisms to regulate behavior will have to be socially established; most importantly, through an internalization of social norms and moral values for the protection of trust and trustworthiness (Ripperger 1998).

Both trustor and trustee enter the trust relation with an established value system and tacit understandings of the socially shared norms and cultural practices, such as interactional routines and role models or standards of economic and social exchange. This renders possible a form of "rule-based trust" (Kramer 1999), which is not based on a conscious calculation of conse-

quences, but on shared understandings regarding the system of rules specifying appropriate behavior (or in the term of Jones and George (1998), based on "value congruence"), which triggers a suspension of doubt and distrust. Such rule-based trust emerges and is sustained by socialization into the structure of normative rules of a given social system. It can acquire a taken-for-granted quality if the members of a social system are highly socialized and experience continued and successful enactment of the rules. This view was endorsed, for example, by Zucker (1986: 54), who explicitly stated that trust was a set of expectations shared by everyone involved in an economic exchange. If trust is regularly experienced by individuals in a social system, it "exists as a social reality, [and] interpersonal trust comes naturally and is not reducible to individual psychology" (Lewis & Weigert 1985b: 976). As Lewis and Weigert argue, trust is essentially "social" and "normative," rather than "individual" and "calculative." In short, trust has to be understood from an institutional perspective as well.

The institutional frameworks that surround a trust relation provide the rules, roles, and routines which can be a basis for trust because they represent shared expectations that give meaning to action (see chapter 3.2 below). To indicate that trust is grounded in social institutions which provide relevant norms and moral values, and thus based socially shared expectations, we will henceforth use the term of *rule-based,* or *institutional trust*. According to Messick and Kramer, rule-based trust resembles a "shallow form of morality," meaning that "the kind of deliberation and thought required to make a decision to trust or be trustworthy is not what psychologists call 'deep' and 'systematic' processing ... we decide very quickly whether to trust or to be trustworthy" (ibid.103).

2.2.5. Feelings and Emotions

Although most researchers agree that trust has a cognitive basis, many maintain that trust is a more complex psychological state, including an affective and emotional dimension as well. Due to the prevalence of cognitive-behavioral accounts, the affective dimension of trust was "historically overlooked" (Lewicki et al. 2006: 997), but researchers have been substantiating claims of its importance, both theoretically and empirically, ever since (Johnson-George & Swap 1982, Lewis & Weigert 1985b, Rempel et al. 1985, McAllister 1995, Jones & George 1998, Williams 2001, Dunn & Schweitzer 2005, Schul et al. 2008, Lount 2010). Concerning the relation between the two dimensions, Bigley and Pearce note that "one of the most contested issues ... relates to whether trust is exclusively the product of individuals' calculative decision making processes or is emotion-based" (1998: 413). In a nutshell, many researchers insist that the cognitive basis of trust is necessary for the understanding of trust phenomena, but in itself not sufficient—one not only "thinks" trust, but also "feels" trust.

To describe affective states, it is important to make a theoretical distinction between *moods* and *emotions* (Schwarz 1990, Clore 1992). The distinguishing feature between them is the in-

tensity of the affective state and the particularity of the affective experience. Emotions are specific affective reactions to particular events. Most importantly, they have an identifiable cause and a clear content (e.g. disgust, anger, joy); they rise quickly, and have a relatively short duration. In contrast, moods can better be defined as low-intensity, diffuse, and enduring affective states which have no salient cause, and also possess less clearly defined content (e.g. being in a good or bad mood). Both mood and emotions indicate how one "feels" about things in daily activities, including interactions with other people. They inform individuals about the nature of the situation in which they are experienced, and they signal states of the world that need to be responded to (Frijda 1988, Damasio 1994).

The influence of mood and emotions on interpersonal behavior has been a prime area of research in social psychology. Researchers have accumulated a large body of evidence portraying the influence of "hot" affective states on memory, judgment, decision making, and the choice of processing strategies across a wide range of content domains (see Schwarz 1990, 1998, Forgas 2002, Schwarz & Clore 2007). Although early psychological research regarded affect mainly as a "biasing" factor to the cognitive process of rational decision making, it is now increasingly accepted as an inseparable aspect of human experience, and one with high informative value to individuals. According to the "feeling-as-information" paradigm (Schwarz & Clore 1983), affective states can serve as an additional source of information while making a judgment. Individuals simplify complex judgmental tasks by asking the question "How do I feel about it?," and use their feelings in a heuristic manner to solve problems. Affective states influence subsequent judgments *directly* when individuals let their judgments be informed by their feelings; they influence judgments *indirectly* when they change the processing strategies of an individual, and thus influence "what comes to mind." With regard to the subjective experience of trust, this suggests that affective states can have both direct and indirect impacts on judgments of trustworthiness and on the choice of a trusting act. In fact, the experience of moods and emotions is sometimes considered a primary aspect of the subjective experience of trust (Jones & George 1998). Yet different authors hold different views on when and how mood and emotions are part of this subjective experience.

To begin with, a trustor might decide on the choice of a trusting act by examining the emotions he has towards a potential trustee in a given trust problem. The experience of positive (e.g. enthusiasm, excitement) or negative (e.g. nervousness, fear, anxiety) emotions can influence the judgment of trustworthiness and the decision to trust (Jones & George 1998, Dunn & Schweitzer 2005). Why would we expect such emotions to be present in the first place? According to "cognitive appraisal theory" (Smith & Ellsworth 1985, Lazarus 1991a, b), emotions result from a sequence of cognitive processes activated whenever the present situation is recognized as having an impact on personal well-being. The mere recognition that something is at stake, and that the outcome of a transaction is relevant to personal well-being, is sufficient

to generate emotions (Lazarus 1991a). As such, emotions should matter in a trust problem as well. A particularly pessimistic interpretation of this fact was delivered by Messick and Kramer (2001), who argue that trust is "bothersome," and may be accompanied by feelings of anxiety, deference, or fear—but generally by negative feelings.

In contrast, Maier (2009: 35ff.) argues that trust is connected to a broad spectrum of different emotional reactions that arise in response to the interpretation of a given trust problem. More specifically, she argues that emotions result from cognitive appraisals in which the current situation is scrutinized for its meaning. This mirrors the idea of interpretation and a "subjective definition of the situation." If a trustor becomes aware of the trust problem, the situation is framed either as an opportunity or a threat, and evaluated in terms of its impact on future personal well-being. Then, expectations of trustworthiness are generated, to which emotional responses automatically develop. Finally, these emotional reactions accumulate into a summary "feeling of rightness"—the anticipated trusting act either feels right, or it does not. The emotions that culminate into the feeling of rightness can be anger, disgust, fear, joy, happiness, love, sadness, and surprise—that is, both negatively and positively valenced emotions. According to Maier, the arousal of each specific emotion is dependent on the context in which the trust problem is embedded. Each emotional reaction reflects a particular set of expectations, the trustor's knowledge about the trustee, and the perceived status of the trust relation. Emotions may also be conflicting or ambivalent. In sum, a trustor "may be cognizant of certain calculated deliberations, intuitions, or expectations, compelling him to trust. Ultimately, he trusts—or not—because it feels right, no matter which factors inform that trust" (Maier 2009: 48).[11]

Emotions may be a result of how a trustor interprets a given trust problem. As such, they are conditional on a preceding interpretive activity ("cognitive appraisal"), in which the trustor uses his knowledge to define the situation and to form expectations. But emotional reactions may spontaneously emerge in a trust problem even without any prior interpretive effort, and may influence perception in form of a mood or an emotion that is not directly related to the immediate interpretation of the trust problem. In this way, affect may influence the process of interpretation and the formation of expectation itself. One source of immediate automatic emotional responses with a measurable impact on judgments of trustworthiness is the recognition of faces. The recognition of human faces is highly automatic and results in the activation of areas in the brain which are associated with the processing of affective stimuli (Whalen et al. 1998, Haxby et al. 2002, Phelps 2006). Empirical neuroscience studies show that the presentation of facial stimuli elicits automatic emotional responses with consequences for the

[11] One can argue that emotional output ("feeling of rightness") and cognitive output ("willingness to be vulnerable") are different sides of the same coin. Although they are empirically hard to separate, we will treat them as analytically distinct.

subsequent evaluation of trustworthiness (Winston et al. 2002, Adolphs 2003). Empirically, Dunn and Schweitzer (2005) show that induced emotions, even when they are unrelated to a specific target, influence trust. They find that negative emotions such as anger, sadness, and guilt reduce judged trustworthiness and the intention to trust a trustee. Likewise, positive emotions such as joy, gratitude, and pride increase judgments of trustworthiness and intentions to trust.

This argument extends directly to the impact of mood on trust. One consistent finding in social psychological research is that global moods have a strong impact on processing strategies (Schwarz 1990, Mellers et al. 1998, Forgas 2002). Positive mood promotes a more global, "top-down" processing style in which individuals tend to rely on existing knowledge structures, while negative moods encourage a more systematic and detailed "bottom-up" processing style, in which individuals rely more on external information (Bless & Fiedler 2006). In line with this, Forgas and East (2008) hypothesize and experimentally demonstrate that a negative mood increases skepticism and decreases the tendency to accept interpersonal communication as truthful. At the same time, people in a negative mood became more accurate at detecting actual deception. Several trust researchers argue that a happy mood globally promotes a more positive perception of others and thus increases trust (Jones & George 1998, Williams 2001). However, Lount (2010) argues and empirically corroborates the idea that a happy mood may also result in less trust. If cues of distrust are situationally available, they are more likely to be used by the happy, "top-down" individuals, resulting in decreased levels of trust. Both Dunn and Schweitzer (2005) and Lount (2010) show that the impact of affective states is neutralized when individuals are made aware of their affective state and its source. All in all, both mood and emotions can influence the choice of a trusting act by exerting a direct or indirect influence on judgments and decision making in a trust problem. They are an ever-present element of the subjective perception of trust.

What is more, if trust is not failed in repeated successful interactions, the trust relationship creates social situations that allow for intense emotional investments. The repeated behavioral expression of trust then reinforces and circulates positive affect (Rempel et al. 1985). Lewis and Weigert (1985b) claim that the development of strong *affective bonds* between actors can extend the cognitive basis of trust. That is, although grounded in cognition, trust can become predominantly a matter of positive affect towards the trustee. This aspect seems to describe most accurately what many authors describe as the "genuine" character of trust, which presumably develops only in the later stages of an interpersonal relation. It is conceived of as a state of positive affect, in which the trustor feels emotionally secure, confident that he will not be exploited, and does not even consider the possibility of opportunistic action on the part of the trustee (Baier 1986, Holmes 1991, McAllister 1995, Becker 1996, Jones 1996, Lahno 2001, 2002). For example, Lahno argues that the particular affective state of trust works like a

perceptive filter, which "has an immediate impact on the beliefs and preferences of a trusting person" (2001: 183).

One plausible explanation for the emergence and prevalence of affective states in mature trust relations is the development and activation of *relational schemata* (Baldwin 1992), which individuals use to frame their social relations.[12] A relational schema is based on the idea that "people develop cognitive structures representing regularities in patterns of interpersonal relatedness" (ibid. 461). Such cognitive structures may be relation-specific or generalized. Importantly, relational schemata also contain typical affective responses and schema-triggered affects (Fiske 1982, Fiske & Pavelchak 1986, Baldwin 1992, Chen et al. 2006). If, by repeated interaction, relational schemata develop which pertain to a particular trust relation and to a particular trustee, their activation can automatically trigger associated moods and emotions (Andersen & Chen 2002). That is to say, the subjective experience of affect in interpersonal trust can, in part, be rooted in the application of stored relational schemata that include affect, and which become activated when a trustor recognizes a trustee to which the relational schema can be applied while facing a trust problem (Huang & Murnighan 2010).[13]

Lastly, researchers frequently point to the negative emotional reactions that emerge when trust is failed (Lewis & Weigert 1985b, Baier 1986, Robinson 1996, Jones & George 1998, Lahno 2001). In an attempt to establish the emotional character of trust, Lahno (2001: 181f.) proposed that trustors adopt a "participant attitude" and see themselves as personally involved and actively engaged in an interaction with the trustee. As a result of this, they become predisposed to "reactive emotions," that is, emotional reactions to the presumed intentions of the trustee. The trustor attributes an intention to the trustee that he himself acknowledges as justified and valuable (e.g. benevolence), and ascribes an implicit normative obligation to the trustee to act appropriately. As we have seen, this aspect of the trustee's "implied moral duty" is common to many trust definitions (Hosmer 1995). Importantly, it lends an emotional charge to the "cold" expectations of trustworthiness. When expectations of trustworthiness are disappointed and the trustee fails trust, the trustor often experiences strong negative emotions, such as disappointment and anger. Robinson (1996) regards the failure of trust as a severe form of "psychological contract breach," which elicits more intense repercussions than unfulfilled expectations, because general beliefs about respect for persons, codes of conduct, and assumptions of good faith and fair interaction—in short, moral values—are violated. This creates a

[12] As Baldwin (1992: 461) points out, relational schemata have been described using other terms, such as *interpersonal schema* (Safran 1990a, b), *working model* (Bowlby 1969), *relationship schema* (Baldwin et al. 1990), *relational model* (Mitchell 1988), and *relational schema* (Planalp 1987). We will here adopt the term "relational schema."

[13] Specific relational schemata can even be applied to contexts and actors in which the particular significant other is *not* present (Andersen & Chen 2002): "Interpersonal cues in a new person, such as the way he or she listens, hold one's gaze, or draws one out, or even his or her smell, gestures, facial features, habits or attitudes, can all serve as applicability-based cues that contribute to the activation of a relevant significant-other representation, along with the associated relational self" (ibid. 623). This is called the principle of *transference*.

sense of wrongdoing, deception, and betrayal, with implications for the relationship in question. Experimental studies reveal that even in simple, anonymous interactions, the violation of expectations results in negative emotional reactions and an impulse to punishment (Fehr & Gächter 2000a, b). Thus, it is reasonable to assume that reactive emotions are most likely a part of the subjective perception of (failed) trust as well.

Unsurprisingly, a large number of authors have provided theoretical or empirical contributions suggesting a distinction between cognitive or affective forms of trust (e.g. Johnson-George & Swap 1982, Lewis & Weigert 1985a, b, Rempel et al. 1985, McAllister 1995, Jones & George 1998). For example, Lewis and Weigert (1985b) propose a distinction between "cognitive trust" and "emotional trust." While cognitive trust is primarily based on "cold" reasoning, and has a low level of affectivity, emotional trust is motivated primarily by positive affect for the trustee, and relies less on its cognitive foundation. Likewise, McAllister (1995: 25) proposes a distinction between one type of trust grounded in cognitive judgments of trustee characteristics (benevolence, integrity, competence, predictability)—which he refers to as "cognition-based trust"—and a second type, founded in affective bonds between individuals, referred to as "affect-based trust." Jones and George (1998) differentiate between "conditional trust," based on knowledge and positive expectations, and "unconditional trust," which is based on shared interpretative schemes and positive affect. All three authors maintain that trust initially emerges from a cognitive foundation and shifts to a more affect-based form only with the continuation of a successful, ongoing trust relationship and the evolution of emotional ties between the interactants.

Similarly, Rempel et al. (1985) distinguish between "predictability," "dependability," and "faith" as unique stages of trust development, where each stage requires an increasing investment in terms of time and emotional commitment. They suggest that the last stage ("faith") is no longer rooted in past experience, but is noncognitive and purely emotional. Thus, faith "reflects an emotional security on the part of individuals, which enables them to go beyond the available evidence and feel, with assurance, that their partner will be responsive and caring despite the vicissitudes of an uncertain future" (Rempel et al. 1985: 97). Note that this is surprisingly close to an affect-based reinterpretation of the idea of suspension. In the opposing perspectives of cognitive versus affective trust, an irrational element of suspension is found in the emotional content bred by "thick" trust. That is, once the affective bonds between individuals become strong and affect-based trust develops, a trustor no longer attends to the cognitive basis of trust.

In sum, affective states are an important aspect of the subjective experience of trust. They provide a trustor with signals concerning the nature and status of an initial or ongoing trust relation in a particular situation. Taking the "leap of faith," a trustor chooses a trusting act because his "cold" expectations are sufficiently favorable and stable, and/or because a corre-

sponding "hot" affective state makes the choice of a trusting act feel right. The automatic arousal of emotions and the presence of incidental moods can influence the judgment of trustworthiness, and emotional reactions or changes in the individual mood state can occur as a consequence of the cognitive appraisals which a trustor executes in a particular trust problem. Depending on the developmental stage of the trust relation, cognitive or affective elements may dominate the choice of a trusting act. But ultimately, trust relations provide the ground for a "mix of feeling and rational thinking" (Lewis & Weigert 1985b: 972), and both dimensions must be assumed to be of equal importance to our theoretical conceptualization of trust.

Recent neuroscience studies suggest that the connection between cognition and affect is much closer than is portrayed in classical psychological and philosophical accounts, where the two are commonly treated as separate. In sharp contrast, cognition and affect seem to be inseparably intertwined at all stages of human experience and development (Reis et al. 2000, Phelps 2006). With respect to the phenomenon of trust, Hardin rightly concludes that "if we wish to separate non-cognitive from cognitive trusting behavior, we will most likely find them thoroughly run together in any kind of data we could imagine collecting" (2002: 69).

According to Reis et al. (2000: 860f.), cognitive expectations provide an important "connecting corridor" between cognition and emotion. Expectations allow the detection of discrepancies between what can be expected, based upon past experience, and the current state of the environment. Many theories of emotion implicitly or explicitly assume that the detection of such a discrepancy is necessary for emotions to arise. The authors conclude that the fulfillment or violation of expectations and the arousal of positive and negative emotions are tied together in every social relationship, and particularly so in trust relations. As pointed out before, one reason for the strong emotional charge of expectations of trustworthiness is that they include a *normative* element (that is, an obligation for benevolence, fairness, honesty, and integrity; a demand for compliance to socially shared norms of trust and trustworthiness). As a result, their violation does not simply result in a "cold" adjustment of expectations, but in a "hot" emotional reaction. Expectations of trustworthiness are often emotionally charged due to their implicit (or explicit) normative content, and the reference to moral values and social norms. This naturally relates trust back to the social systems surrounding a particular trust relation, and to the cultural and institutional structures in which the trust problem is embedded.

2.3. Conceptual Boundaries

2.3.1. Familiarity and Confidence

"Familiarity, confidence, and trust are different modes of asserting expectations—different types, as it were, of self-assurance" (Luhmann 1988: 99). This proposition of Luhmann, wide-

ly accepted and regularly adopted into theoretical frameworks of trust, will serve as a starting point for a discussion of the conceptual boundaries of the concept of trust. As it is, the relationship between the three constructs is challenging. The concept of *familiarity* points to a central factor of human experience: it describes the certain acceptance of a socially constructed reality as *the* unique reality, which is not questioned in terms of its consistence or validity. In consequence, familiarity implicitly precedes any action as an underlying assumption of *taken-for-grantedness* in the "natural attitude" towards the life-world (Berger & Luckmann 1966, Schütz & Luckmann 1973). Familiarity means that situations which would otherwise be considered problematic can be automatically recognized as typical through the use of learned interpretive schemes. This frees up cognitive resources to plan and engage in future-oriented actions. In this sense, familiarity must be regarded as a precondition to trust, and the conditions of familiarity and its limits must be considered when thinking about trust (Luhmann 1988). Luhmann asserts that familiarity is directed towards the past and things already known, while trust points toward the future and unknown things (namely the trustee's intention and behavioral response). In essence, trust has to be achieved within in a familiar world, and trust and familiarity are "complementary ways of absorbing complexity, and are linked to one another, in the same way as past and future are linked" (Luhmann 1979: 20)

However, Luhmann's analytic distinction is problematic. By definition, it precludes the possibility of grounding trust in the recognition of the "typical" and in socially shared interpretive schemes, that is, in rule-based forms of trust, which rest upon familiarity with the cultural and normative content of the social systems and the institutions surrounding a trust relation. Luhmann argues that trust, in contrast to familiarity, "risks defining the future" (ibid.). Yet a recognition of the familiar implicitly does the same, because it presupposes an idealization of an ongoing, unproblematic present. Thus, familiarization is also future-oriented (Möllering 2006a).[14] On the other hand, trust also bears an orientation to the past, because it implicitly rests on the fact that trustworthiness has not been failed "so far," and it extends this fact, as a taken-for-granted background assumption, to the present and the future (Endress 2001). The boundary between trust and familiarity is thus fuzzy and less clearly defined than suggested by the terminological framework of Luhmann. Möllering (2006a) argues that the process of familiarization, akin to suspension, must be regarded as a core element of trust, rather than as a "fringe consideration," or an otherwise distinct concept. This is important, because trust can approximate familiarity when the situation is structured to such an extent that recognition of the typical is sufficient to induce trust. For example, Misztal (2001) discusses trust as an outcome of situational normality, based on "familiarity, common faith, and values" (ibid. 322). In

[14] Möllering (2006a) uses the term *familiarization* to describe an important aspect of interpretation. Familiarization describes recognition of the typical, and it includes the possibility that unfamiliar things, by recognition of their similarity to the typical, may be "brought in" to familiarity without interruption of the ongoing routine of pattern recognition.

a similar argument, McKnight et al. (1998) include beliefs of *situational normality* and *structural assurance* into their model of trust development, arguing that both antecedent factors have a direct impact on the trustor's expectation of trustworthiness and his trusting intentions.[15]

Luhmann's second analytic distinction between trust and confidence is equally problematic. It is a distinction made upon an assumption on the level of subjective experience: both trust and confidence refer to expectations of future contingencies that may disappoint. Yet the term c*onfidence* is applicable whenever an actor subjectively does not take into account the potential for damage and harm (Luhmann 1988). Luhmann refers to confidence as the "normal case." Individuals do not expect their everyday routine to break down. They would, in fact, not be capable of acting, if such a state of permanent uncertainty prevailed. This links confidence directly to familiarity and to the recognition of the typical. While familiarization describes the background process, confidence specifies its subjective experience. As such, confidence can be understood as "a kind of blind trust" (Gambetta 1988a: 224), in which alternatives are not taken into account and vulnerability is suppressed from actual perception. If such confident expectations are disappointed, the reasons for the failure must be found in external conditions. On the other hand, in the case of trust, an actor must be cognizant of the possibility of harm and vulnerability. If trust fails, the trustor will need to attribute the failure to his mistaken choice of action. Importantly, it is only when the trustor is aware of the potential of damage, of his vulnerability, and of the risks involved, that we can speak of trust (Luhmann 1988: 98). But as we have seen, conceptualizations of trust widely diverge on the question of whether trust is a deliberate and intentional phenomenon, and a product of the conscious calculation of risks. Luhmann's distinction between confidence and trust, based on the criterion of subjective experience, categorically excludes the possibility of any form of blind trust. Yet such blindness is regarded as a characteristic feature of trust by other researchers. His distinction is problematic because trust problems can be solved *in confidence* as well. That is, instances of "confident" trust (for example, when rule-based or affect-based) do not necessarily rely on a conscious and deliberate experience of risks and the consideration of potential harm and vulnerability. As Luhmann notes, a situation of confidence may turn into a situation of trust when the inherent risks become perceptible and the alternative of avoidance (that is, distrust) is taken into account, or *vice versa*. The problem of how and when a situation of confidence turns into a situation of trust is, as he notes, intricate, and it points to the centrality of the process of interpretation to our understanding of trust.

[15] According to McKnight et al. (1998), *situational normality* beliefs indicate the appearance that things are normal and in proper order, which facilitates successful interaction. This refers to the concept of familiarity, and includes the actor's acquaintance with social roles. *Structural assurance* beliefs, on the other hand, involve the opinion that trustworthiness will be guaranteed because contextual conditions, such as promises, contract, and regulations are in place. The authors subsume these two factors under the label "institution-based trust."

Owing to the fact that trust and confidence are convertible, Luhmann observes that "the relation between confidence and trust becomes a highly complex research issue" (1988: 98), and admits that, "belonging to the same family of self-assurances, familiarity, confidence, and trust seem *to depend on each other* and are, at the same time, *capable of replacing each other* to a certain extent" (ibid. 101, emphasis added). In developing a broad theory of trust, we will have to pin down the relation between these concepts, and explain which factors, both internal and external, facilitate the transition between these different states of subjective experience. In short, what is needed is a theory of interpretation that is capable of causally explaining the emergence of either type of "self-assurance." As will be argued in the course of this work, a major factor that determines the subjective experience of trust is the information processing state of the cognitive system. In this perspective confidence and trust, which face the same structural prerequisites, differ mainly with respect to the degree of rationality and elaboration involved in dealing with the trust problem.

2.3.2. Self-Trust

The concept of trust can also be applied in a self-referential way. More pointedly, *self-trust* denotes cases of trust where subject and object of trust are the same. This literally means to "trust in one's own identity." Govier (1993: 105f.) claims that self-trust is completely analogous to interpersonal trust, in that it includes the same defining features: (1) positive expectations about one's own motivations and competence, (2) a self-attribution of personal integrity, (3) a willingness to rely on oneself and also to accept risks from one's own decisions, and (4) a disposition to see oneself in a positive light. But what would constitute a trust problem with respect to one's own identity? In our terminology, a problem of self-trust would portray an aspect of uncertainty with respect to one's own future preferences, motivations or competence; it thus recognizes the possibility of preference change or a deterioration of skills. Essentially, it would amount to the question whether some investitive action can be justified by one's own anticipated future preferences and motivation. But the concept of action always includes a motivation in the form of the desired ends, induced by the current preferences, and the purposively chosen means which cater to their fulfillment. If an action is chosen for the realization of some end, this implicitly rests on the assumption that the preferences which have motivated the action will remain stable, and that utility can be realized accordingly in the future. If preferences were completely transitory and random, an actor would be incapable of action. Uncertainty with regard to future preferences is most unlikely: we still know ourselves the best. A suitable way of making sense of "self-trust," then, is to regard it as a type of "confidence" in one's own competence, skill or character qualities.

Thus, the term self-trust is closer to, and should be consistently replaced by, the concept of self-esteem and self-"confidence."[16] These refer to the balance of positive and negative conceptions about oneself and the certainty of the clarity of such self-conceptions (Banaji & Prentice 1994). Self-esteem and self-confidence are positively related to trust: high self-esteem increases the readiness to engage in trusting behavior because individuals with high self-esteem tend to subjectively experience an augmented feeling of control over the environment, which is conducive to the acceptance of risks (Luhmann 1988: 82). The mechanism behind this effect is referred to as the "illusion of control" bias (McKnight et al. 1998, Goldberg et al. 2005). McKnight et al. (1998) add illusion of control to a set of cognitive processes that interact with other antecedent factors to elevate expectations of trustworthiness. It moderates the effect of general dispositions to trust (that is, it increases the "illusion" that generalized expectancies apply to particular instances), of categorization processes such as stereotyping (that is, it builds confidence that applied categories are correct), and of structural assurance beliefs (that is, it reassures the conviction that structural safeguards are secure and effectively procure trustworthiness). In short, self-esteem is not a direct antecedent to trust, but rather an indirect antecedent which influences or "biases" the degree of certainty that a trustor can have with respect to his expectations. It is relevant to a conception of trust inasmuch as it indirectly influences how individuals deal with others during an interaction and how they approach the environment in general.

2.3.3. System Trust

System trust, in contrast to interpersonal trust, refers to abstract institutions or social systems as *objects* of trust. Luhmann (1979: 48f.) introduces the concept of system trust by analyzing the monetary system and its stability, which, as Luhmann argues, is maintained by the trust of the participating individuals in the functioning of the system as a whole. System trust is created and sustained by the continual, ongoing, confirmatory experience of the system's functioning. In contrast to interpersonal trust, system trust does not concern social uncertainty with respect to another individual's action, but the global characteristics of an institution: its primary goals, its legitimacy, structure, and operation, and the effectiveness of the sanction mechanisms which structure and control interaction in social settings (Endress 2002: 59). This notion of trust in the system, especially at the macrolevel of society, has been a prominent topic of trust research (Fukuyama 1995, Putnam 1995, Sztompka 1999, Cook 2001). Trust in the reliability, effectiveness and legitimacy of money, law, and other cultural symbols warrant their smooth functioning and constant reproduction, and the absence of system trust facilitates

[16] Confidence is not used in Luhmann's sense here. Govier (1993) uses the terms self-confidence and self-trust interchangeably. Her conceptual ambiguity is revealed most clearly when discussing the *absence* of self-trust, or distrust in oneself: she compares this to a "lack of confidence" (ibid. 108) and "extreme self-doubt" (ibid.). These conceptual slippages go to show that what she is really addressing is self-esteem, or self-confidence. In her work, the analytical difference between self-trust and self-esteem remains veiled.

the deterioration, decline, and ultimately the disruption of a social system (Lewis & Weigert 1985b).

It is important to distinguish the role that institutions can take as a basis of expectation formation in a trust problem from their role as an object of trust. We have referred to the former case as rule-based, institutional trust, and we will denote the latter case as system trust. In this sense, institutional trust is a manifestation of system trust in a particular interaction. Luhmann argues that system trust ultimately depends on a form of generalized trust, or "trust in trust" (1979: 66f.). That is, system trust rests on the assumption that other actors in the social system do equally trust in it. According to Luhmann, system trust is impersonal, diffuse, rests on generalizations, and—in contrast to interpersonal trust—is marked by a low degree of emotional investment and affectivity.[17] An actor who participates in a social system to which he maintains a high level of system trust assumes that the actions of other actors in the system are effectively regulated and structured by the institutionalized norms, rules, and procedures. Most importantly, system trust includes the expectation that norm violations are effectively sanctioned. These background assumptions, shaped by system trust, lay the ground for institutional trust to emerge. Luhmann proposes that the basis of system trust is the appearance of normality (1979: 22, Lewis & Weigert 1985a: 463). This indicates that there is a link joining system trust to the concepts of familiarity and confidence: system trust situationally manifests itself in the form of taken-for-granted background assumptions, that is, in familiarity with and confidence in the functioning of the system and its legitimate primary goals, rules, and sanctioning potential. At the same time, system trust is the basis for institutional trust. If a trustor does not believe that a social institution is effectively regulating and sanctioning the behavior of others, then there is no sense in grounding expectations of trustworthiness in assumed norm-compliance. While institutional trust concerns the concrete interpretation of the institutional rules with respect to the trustee's action, system trust concerns assumptions concerning their general validity, applicability, and enforcement.

At the same time, successful interpersonal trust relations that were structured by institutional trust, and in which trust was not failed, strengthen system trust. This exemplifies how dyadic trust relations can be regarded as "building blocks" of larger systems of trust (Coleman 1990: 188). For example, Giddens (1990: 79f.) argues that trust in the medical system is developed to a large extent through experiences with doctors and medical professionals who represent and "embody" the institutions of medicine, and to whom a patient develops a concrete interpersonal trust relation. In modern societies, a dense network of such institutional intermediar-

[17] The last proposition is problematic. As we have seen, trust between two actors may also be rather non-affective, calculative, or rest on generalized expectations. On the other hand, system trust clearly has an affective dimension (consider, for example, the intense emotional reactions that might arise in response to a violation of an oath of political office, or to corruption).

ies of trust controls the agency of trust between the microlevel and individual actions and the macrolevel of system trust (Zucker 1986, Shapiro 1987, Coleman 1990, Giddens 1990, Mishra 1996). These intermediary institutions—for example courts, product testing agencies, and doctors' surgeries—function as generators of both (rule-based) interpersonal trust and system trust, by providing generalized expectations for trust relations among a number of otherwise anonymous actors.

With high system trust, institutions can serve as "carriers" of trust. Trust in the system is regarded as a public-good resource which facilitates the production of social capital (Ripperger 1998: 164ff.). Since trust relations are embedded into a social environment, system trust is highly relevant for interpersonal trust: a low level of system trust makes interpersonal trust "more risky" (Lewis & Weigert 1985a: 463). Simply put, low system trust reduces the basis of any institutional-based form of trust. Trustors then cannot base trust on structural assurance and on the institutions which normally control and regulate trust and trustworthy action. On the other hand, high levels of system trust can have a positive effect: interpersonal trust becomes less risky because a trustworthy response based, for example, on institutional rules can be *confidently* expected. In sum, "trust occurs within a framework of interaction which is influenced by both personality and social system, and cannot be exclusively associated with either" (Luhmann 1979: 6). As a basis for institutional trust, system trust is an important factor that shapes interactions and trust relations on the microlevel.

2.3.4. Distrust

The concept of distrust has received less scholarly attention than trust, but an ever-growing part of the trust literature focuses on its relation to trust and the problems of its theoretical conceptualization (Worchel 1979, Sitkin & Roth 1993, Bies & Tripp 1996, Lewicki et al. 1998, McKnight & Chervany 2001, Lewicki et al. 2006, Schul et al. 2008, Keyton & Smith 2009). Although the link between trust and distrust seems to be straightforward, a closer look reveals that their relation is intricate and warrants closer inspection.

Luhmann (1979: 71ff.) regards trust and distrust as "functional alternatives" among which the trustor necessarily has to choose, and characterizes distrust as "positive expectation of injurious action" (ibid. 72). Barber (1983) defines distrust as "rationally based expectations that technically competent role performance and/or fiduciary obligation and responsibility will *not* be forthcoming" (ibid. 166): by adding the word "not" to his definition of trust, he creates a definition of distrust. Deutsch (1958) uses the term "suspicion" to denote a state in which a trustor "perceives that he is an object of malevolent behavior" (ibid. 267). Such a state has consequences for the trustor's motivation to engage in trusting behavior. Distrust manifests itself in the choice of the safe alternative instead of the trusting act. The trustor does not transfer control over events or resources to the trustee, and forfeits the potential gains that could be

achieved with trust and trustworthy response. He does not become vulnerable to the actions of the trustee, or confront social uncertainty with respect to the trustee's choice. In short, distrust objectively minimizes vulnerability and social uncertainty, and it can be regarded as the "minimax" solution to a trust problem (Heimer 2001). It has been associated with increased monitoring and defense behavior (Schul et al. 2008), with refusal of cooperation, with reliance on contracts and formal agreements, and with neither party accepting the other's influence or granting autonomy to the other (McKnight & Chervany 2001).

From such a perspective, trust and distrust range on a single dimension and are mutually exclusive (Worchel 1979); they reflect opposite levels of the same underlying construct (McKnight & Chervany 2001). This "bipolar" (Lewicki et al. 1998) perspective on trust and distrust emphasizes that, as trust *decreases*, distrust *increases*. Accounts in the cognitive behavioral tradition support such a conception, in which distrust is regarded as a low level of trust, resulting in the choice of the safe alternative. Recall that Gambetta (1988) locates expectations of trustworthiness on a probabilistic distribution with values between complete distrust (0) and complete trust (1), with a region of indifference (0.5) in the middle. The choice of a trusting act requires that an expectation exceeds a subjective threshold value. If expectations do not exceed the threshold value, distrust and the choice of the safe alternative will prevail. In essence, distrust simply means that expectations of trustworthiness are *unfavorable*, hence the trustor's trusting intention and willingness to become vulnerable are low, and do *not* support the behavioral taking of risks. In consequence, the antecedent conditions to distrust are opposite to those of trust. Distrust can be a result of perceived value *in*congruence (Sitkin & Roth 1993, Jones & George 1998), it can be triggered by *un*favorable assessments of trustee characteristics (including competence, benevolence, integrity, and predictability), and it is fostered by the absence of structural assurance and a lack of perceived situational normality (McKnight & Chervany 2001).

With respect to the affective dimension, distrust is commonly related to negatively valenced emotions like doubt, wariness, caution, defensiveness, anger, fear, hate, and feelings of betrayal and vulnerability (Lewicki et al. 1998, Keyton & Smith 2009). Thinking in terms of "cognitive appraisals," the arousal of these emotions mirrors the formation of negative and unfavorable expectations during the process of interpretation. The resulting emotions signal a potential threat to the trustor. Importantly, distrust is often regarded as a reflective phenomenon, relying on systematic and elaborated processing strategies which allow the trustor to take rational decisions towards necessary protective measures (Luhmann 1979, Lewicki et al. 1998, Endress 2002: 76). In this line, Schul et al. (2008) directly connect trust and distrust to information-processing strategies. In their conceptualization, trust and distrust span a continuum of mental states, which contain, as extreme end points, (1) the use of routine strategies and a feeling of security in the case of trust, and (2) the nonroutine use of elaborated processing

strategies, accompanied by a feeling of doubt, in the state of distrust. While trust is regarded as a "default" state, distrust prevails whenever the environment signals that something is not normal, that is, when situational normality is disturbed. Likewise, Luhmann argues that distrust develops "through the sudden appearance of inconsistencies" (1979: 73), and triggers a "need" for more information (this suggests an interesting connection between distrust and disruptions of familiarity, see chapter 4). The resulting elaborate and protective strategies that the trustor uses "give distrust that emotionally tense and often frantic character which distinguishes it from trust" (ibid. 71).

On the other hand, some researchers have drawn scholarly attention to the possibility that trust and distrust may be separate, but linked (Sitkin & Roth 1993, Lewicki et al. 1998, McKnight & Chervany 2001, Lewicki et al. 2006, Keyton & Smith 2009). In this perspective, trust and distrust are assumed to be independent constructs. The two-dimensional approach is grounded in two observations: Firstly, from a structural standpoint, relationships are "multifaceted and multiplex" (Lewicki et al. 1998: 442), offering a host of different contexts and situations in which two actors face trust problems. Thus, in the context of repeated interaction and social embeddedness, actors maintain different trust relations to each other simultaneously. Trust is content-specific: we may trust someone with respect to X, but not with respect to Y. Consequentially, within the same relationship, both trust and distrust may "peacefully coexist" (Lewicki 2006: 192), since they present solutions to different specific trust problems. Secondly, proponents of a two-dimensional approach emphasize that psychological research examines the possibility that positively and negatively valenced emotions are *not* simple opposites. While emotions have traditionally been regarded as bipolar and mutually exclusive (Russell & Carroll 1999), this view has been challenged by the idea that positive and negative affect are independent, differing even on a neural basis (Cacioppo & Berntson 1994, Larsen et al. 2001).

This suggests that trust and distrust—often portrayed as entailing opposed affective states—may also be independent. Although trust and distrust represent "certain expectations" (Lewicki et al 1998: 444), the content of these expectations may be different: while trust appreciates beneficial conduct from the trustee, distrust points toward the apprehension of harm and defection. That is, a low level of trust may not generally equate to a high level of distrust, and *vice versa*. Lewicki et al (1998), in favor of a multidimensional approach, argue that relationships are dynamic and that trust and distrust are sustained at specific levels akin to a "quasi stationary equilibrium." During interaction, the "operational levels" of trust and distrust move on both dimensions, and change the nature of the "relationship orientation."[18] This sug-

[18] The ideal-type condition of "high trust/low distrust" is, for example, characterized by having "no reason to suspect the other," high "value-congruence," and "strong positive affect," while distrust is characterized as "cautious" and "guarded" behavior, in which a trustor follows the principle of "trust but verify," "fears" undesirable events, and attributes "sinister intentions" to the trustee (Lewicki et al. 1998: 446f.).

gests to the reader that trust and distrust may exist at the same point in time when the trustor faces a trust problem. But how can we understand this "simultaneity"? Is the conceptual independence of the two constructs theoretically justifiable?

Let us step back and examine the issue from a general perspective. The conceptual difficulty with the two-dimensional approach to trust and distrust arises, as is argued in the following, mainly because objective structural conditions and subjective experience are confused. With a clear distinction of the levels of analysis, the ostensible contradictions disappear, and it is easy to see that the multidimensional approach does neither have a solid and logical conceptual foundation, nor add to our understanding of the trust phenomenon.

First, on the level of objective structure, a trustor must *either* choose a trusting act and become vulnerable, *or* choose the alternative of distrust and *not* become vulnerable, for a specific trust problem and a specific content X. Given that actors are socially embedded and interact repeatedly, consecutive trust problems may be solved differently. But many day-to-day trust problems resemble one-shot situations and do not allow for the "simultaneous" existence of trust and distrust. It is problematic to invoke the notion of "simultaneity" when really what is meant is that every trust problem may have a different solution. The domain-specificness of trust in ongoing relations was introduced earlier through the competence dimension of trustee characteristics, and it has already been included in our basic definition of a trust relation, in which "A trusts B with respect to X." Different solutions to different trust problems in repeated interaction do not collide with a unidimensional view of trust and distrust. Of course, the experience of a failure of trust will change future expectations of trustworthiness, and therefore cater to the potential uncertainty a trustor may subjectively experience. But the domain-specificness of trust does not logically imply its conceptual independence from distrust.

Second, in a psychological sense, and on the level of subjective experience, the simultaneous existence of trust and distrust may be better described as the arousal of ambivalent affective states. *Ambivalence* denotes a state in which positive and negative affective reactions towards a target collide (Priester & Petty 1996).[19] Recall that conflicting and ambivalent emotions may very well be a result of "cognitive appraisals" when a trust problem is subjectively defined. Deutsch has already noted that "when the fulfillment of trust is not certain, the individual will be exposed to conflicting tendencies to engage in and to avoid engaging in trusting behavior" (1958: 268). But as Maier (2009) has argued, trust ultimately either feels right, or it does not. In fact, ambivalent affective states are short-lived and transitory, and individuals, for the majority of their everyday lives, can effectively reduce affective ambivalence to a state with a clear valence (Larsen et al. 2001). This argument also reverberates in Möllering's (2001) idea

[19] For example, fear *and* excitement may occur in a new and unknown situation.

of suspension, which addresses those cases where the resolution of subjectively experienced ambivalence is conducive to the successful build-up of trust.

All in all, while the unidimensional view seems to be appropriate on the descriptive level of objective structure and behavioral outcomes, a view that admits the possibility of ambivalent emotional states seems to be appropriate on the level of subjective perception. Ambivalent emotions can be a result of expectations which are not favorable, and reside in the "region of indifference" or below. However, the experience of ambivalent emotions does not logically imply the conceptual independence of trust and distrust. It merely indicates and signals to the trustee that expectations are close to unfavorable, nearing the threshold at which distrust will be exercised.

Lastly, the two-dimensional position does add to the "potpourri" of trust definitions, as it implies illogical types, such as high trust paired with high distrust (Schoorman et al. 2007). The definitions and constructs which have been developed to describe distrust are identical to those of trust, but merely formulated as opposites (McKnight & Chervany 2001).[20] In short, there is no theoretical advantage gained and no explanative value added by treating them as "separate, but linked" constructs. Examining theoretical and empirical work, Schoorman et al. (2007) conclude that there is "no credible evidence that a concept of distrust which is conceptually different from trust is theoretically or empirically viable" (ibid. 350). Because a unidimensional perspective is theoretically more sparse, more tractable, and merges more easily with psychological accounts of trust versus distrust and their relation to information-processing strategies (e.g. Schul et al. 2008), we will maintain it in the following. The preceding discussion nonetheless shows how important it is to have a clear distinction between objective structural conditions, potentially diverging subjective experiences, and the theoretical concepts developed to incorporate both of these into our understanding of the phenomenon of interpersonal trust.

2.4. From Structure to Experience

The preceding analysis of the objective structure of trust and its subjective experience has delineated core elements of the trust phenomenon and served to introduce important theoretical concepts used in the trust literature. Taken together, publications from different research traditions and paradigms seem to converge on some fundamental points. Most importantly, there is broad consensus on the objective structure of trust and on the conditions which give rise to a

[20] "Most trust theorists now agree that trust and distrust are separate constructs that are the opposites of each other" (McKnight & Chervany 2001: 42). The authors develop separate conceptual models for each construct, which are completely identical but contain opposite elements, such as "distrusting beliefs" of competence, benevolence, integrity, and predictability, "institution-based distrust," "distrusting intentions," and "distrust-related behavior."

trust problem. A trust problem is marked by a particular incentive structure and asymmetric, imperfect information; the choice of a trusting act entails objective vulnerability and results in social uncertainty with respect to the trustee's actions. This combination of opportunity and vulnerability is the *sine qua non* of a trust problem.

The crux of trust research, however, is the question of how trustors *subjectively* handle the objective structure of trust; how they make sense of the structural conditions they face. Most trust researchers agree that trust is a special way of dealing with social uncertainty and imperfect information. Trust points to the particular nature of the expectations involved and a particular way that they emerge and form, and it additionally indicates affective processes that accompany and shape the subjective experience of trust. When choosing to trust, a trustor bypasses social uncertainty and convinces himself that the choice of a trusting act is justified. But the propositions made by researchers to illustrate this idea are controversial. Formulated very generally, the question that provokes diverging views on trust is: how does the objective structure of trust translate into subjective experience? How do trustors deal with the social uncertainty present in a trust problem? How is the choice of a trusting act "decided" on? In essence, the concept of trust is blurry, because researchers are unclear about the role of interpretation; they disagree on how trustors subjectively perceive and deal with the trust problem. The process of a subjective "definition of the situation," although crucial to an understanding of the trust phenomenon, is often mentioned in passing only, or it is taken for granted and rarely dealt with explicitly. In short, the conversion from structure to experience, and the cognitive mechanisms involved in doing so, present a "missing link" in trust theory.

Consequently in the transition from structure to experience, the concept of trust often loses its clarity and precision. We have encountered this problem, for example, when dealing with the question of vulnerability. As an objective fact, vulnerability is undoubtedly present in all trust problems. But authors disagree on whether it is subjectively experienced by a trustor, or not. When trust is defined as "willingness to be vulnerable," it involves a conscious perception of vulnerability, by definition. But this is contrary to theoretical accounts which instead relate trust to the suppression of vulnerability; to a sort of innate security and the absence of doubt. In going from structure to experience, the aspect of vulnerability can be retained or rejected, and the way we deal with it depends on our assumptions concerning the process of interpretation. Both notions seem equally plausible, and to exclude either possibility *a priori* would be unnecessary restrictive to a broad conceptualization of interpersonal trust. To answer the question of how and when we can expect vulnerability to be part of subjective experience, it is essential to be more precise about the "missing link" between structure and experience.

We find the same problem when we address the relation between objective uncertainty and the subjective perception of risk or ambiguity. From a cognitive behavioral perspective, the perception of subjective risk is required, built upon the retrieval of imperfect information, and

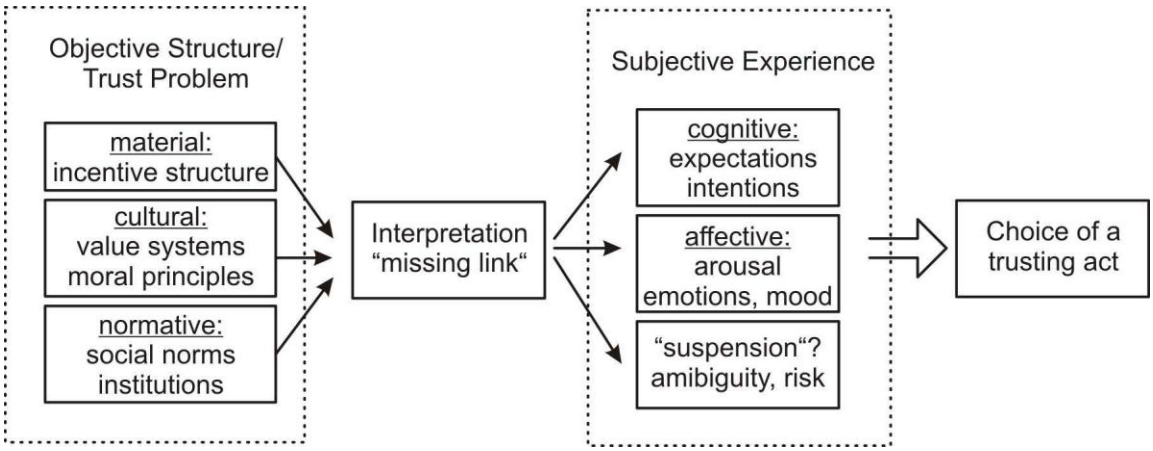
expressed in unambiguous expectations. This notion is rather narrow in comparison to the oft-purported "leap of faith," which enables the trustor to cross "the gorge of the unknowable from the land of interpretation into the land of expectation" (Möllering 2001: 412). Although many authors agree, with Lewis and Weigert (1985b), that trust allows for a "cognitive leap" away from social uncertainty and into stable expectations, a precise formulation of this process is rarely offered. It can be regarded as the stabilization of ambiguity into risk, or an extrapolation of either form of subjective uncertainty into subjective certainty. Clearly, the experience of trust, and hence our conceptions of it, will differ between these possible readings of suspension. On top of that, if suspension is part of the trust phenomenon, it adds a further, nonrational element to a purely cognitive account of trust; it extends our understanding of the concept to something "beyond" the retrieval of knowledge. In going from structure to experience, objective uncertainty may be retained, reduced, or even suspended, and the answer depends on our assumptions concerning the process of interpretation. Möllering (2001) consequently, and rightly, argues that the interpretation and the subjective definition of the situation must be regarded as key processes to trust.

A direct reference to the process of the definition of the situation, and a plea for its importance in our understanding of trust, was made by Jones and George (1998), who develop their ideas with recourse to the paradigm of symbolic interactionism (Mead 1967, Blumer 1969). In their contribution, the notion of "unconditional trust" expresses the idea that trust is built upon a joint social definition of the situation, based on shared interpretive schemes which contain values, attitudes, and affect, and which serve as structuring devices for actions in the trust problem, facilitating the suspension of perceived risks and doubt (Jones & George 1998: 535f.). However, according to the authors, this requires that the actors involved have already developed a strong confidence in each other's values and trustworthiness through repeated interaction, and that they hold favorable assessments of each other's characteristics, which, taken together facilitate the experience of positive affect. "Conditional trust," on the other hand, is purely based on knowledge and a favorable assessment of trustworthiness. It does not go along with an equal amount of suspension and affectivity—the actors only engage in what Jones and George term a "pretense of suspension." The differentiation between these two types of trust closely resembles the many contributions in which "cognition-based" and "affect-based" forms of trust are treated as separate ideal types. Again, the "missing link" of interpretation seems to be of prime importance. In going from structure to experience, trustors can use different strategies to deal with a trust problem, resulting in conceptually different "types" of trust. Clearly, the emergence of these "types" rests on our assumptions concerning the process of interpretation. A broad theory of trust must predict whether, how, and when trustors will use different strategies to solve a trust problem, and it should demarcate the types of trust that will consequently appear.

Another instance that highlights the importance of the definition of the situation for the subjective experience of trust is the question of affect. As we have seen, short-lived emotions arise during active interpretive processes ("cognitive appraisal"). The situation is scrutinized for its impact on well-being and framed for its valence, either as an opportunity or a threat. Expectations are generated, and, finally emotions surface as a response to these interpretive efforts. These affective reactions include a broad range of emotions, which mirror how favorable the expectations of trustworthiness are. But researchers have also related trust to relatively stable and long-lived positive mood states which function as perceptive filters and directly impact expectation formation and preferences. In this perspective, a "trustful" affective state precedes the conscious formation of expectations, straightforwardly biasing them. What is more, affect-based forms of trust are sometimes seen to emerge exclusively of any deliberate efforts to analyze the situation. Trustors choose a trusting act based on positive feelings, without further scrutinizing the trust problem, and without further thinking about the risks involved. That is, in going from structure to experience, affect can be a by-product of a controlled reasoning process, or it can be seen to operate preconsciously, activated for example by an automatic use and application of interpretive schemes or relational schemata. Affect may precede, replace, and bias cognitive expectations and, when used as information in its own right, can serve as a quick-step heuristic to interpersonal trust. Whether, how, and when affect is used heuristically to inform the choice of a trusting act, again depends on our understanding and conceptualization of the process of interpretation.

Lastly, the normative dimension of trust highlights the importance of the social systems into which a trust relation is embedded. Cultural and normative systems deliver the moral values, cultural practices, and social norms that structure interaction in general, and interactions in a trust problem in particular. If norm-compliance is regarded as a viable factor in determining trustworthiness expectations, then interpretation is the implicit propellant behind the emergence of rule-based forms of trust. As we have seen, rule-based forms of trust are often regarded as noncalculative forms of "shallow trust" (Messick & Kramer 2001), which do not require deep, systematic processing efforts. They rest on shared understandings regarding the system of rules specifying appropriate behavior, instead of being the result of a rational calculation of consequences. Implicitly, what is taken for granted here is that actors can easily identify situations in which "shallow trust" is feasible, and adjust their strategies accordingly. In other words, in going from structure to experience, trust may approximate a rational choice built upon the calculation of expected utility; it may also emerge based on the "shallow" use of rules and other heuristics, such as norm or relational schemata. Yet again, the "missing link" of interpretation and our assumptions concerning how people make sense of their environment and use the available knowledge are decisive in describing whether, how, and when we can predict trust to be more calculative-, or rule-based (see figure 6).

Figure 6: Interpretation – the "missing link" in trust research



Overall, when thinking about trust, the process of the subjective definition of the situation seems to be key in specifying the phenomenological foundation, the "mindset" of trust, and the associated subjective experiences. The decisive question that makes trust so difficult to grasp is that of how objective structural conditions translate into subjective experience. In order to understand trust, we must sharpen our understanding of the "missing link" of interpretation. Controversial accounts of trust arise because different authors specify different aspects of our social reality as being relevant to the problem of interpersonal trust, and, in doing so, they implicitly make divergent assumptions concerning the underlying process of interpretation. This extant state of affairs gives the impetus for the present work, which seeks to open the "black box" of interpretation in order to advance our understanding of the trust phenomenon.

The argument that will be developed in this remainder of the book can be summarized thus: to understand why there are such dramatic differences between our understanding of the objective structure of trust and its subjective experience, it is necessary to connect the process of interpretation to the concept of *adaptive rationality*. Recent social psychological and neurological research suggests that information can be processed at different degrees of elaboration and detail. In light of the discrepancies presented above, the assumption that actors are sometimes "more rational" and sometimes "less rational" seems to be a promising approach to remedying the current contradictions. If we think of actors as being capable of a flexible degree of rationality, then we can understand, for example, why vulnerability and risk are perceived at some times, and why actors suspend them at other times. Broadly speaking, adaptive rationality must be regarded as an important underlying dimension of the trust concept, determining the different strategies which actors use to solve a trust problem. Interpretation is then the "motor" behind the adjustment of the degree of rationality involved. Although many authors implicitly refer to adaptive rationality when specifying different types of trust, it has not been systematically incorporated into current theoretical frameworks, nor given the central status it deserves. What is more, a precise formal model that allows for a tractable conceptual-

ization of adaptive rationality as an endogenous factor of interpretation has not yet been put forward. The present work seeks to advance our understanding of trust by linking it directly to interpretation, adaptive rationality, and the choice of a trusting act as a behavioral outcome. Ultimately, it seeks to contribute to the development of a unifying framework in which clear theoretical predictions about the different forms, types, and nuances of trust can be made.

3. Origins and Explanations: An Interdisciplinary Approach

"Experience molds the psychology of trust" (Hardin 1993: 508).

Having defined the major components of the trust concept—its objective structure and subjective experience—we can now take a closer look at the theoretical approaches and major research paradigms of trust research. The overarching idea in this discussion is to provide an interdisciplinary perspective and to outline the commonality, mutuality, and similarities that allow the existing theory to be integrated into a broader theoretical framework. In short, the psychological, sociological, and economic disciplines, by each emphasizing different aspects of the trust phenomenon, provide a unique perspective on trust that deserves proper and detailed presentation. In developing an interdisciplinary perspective, the aim is not to discuss the pros and cons of each paradigm and to finally opt for one of them, or rule out another. Instead, we want to carve out the reasons behind their incompatibility and the lack of cross-disciplinary fertilization. As we will see, the most essential factor which prevents the development of a more comprehensive and integrative perspective is, across disciplines, an insufficient consideration of the aspect of adaptive rationality.

The discipline of psychology explores the subjective experience of trust in the form of internal cognitions and affect along with the conditions of its emergence, maintenance, and disruption. This entails an analysis of the developmental aspects of trust, both in terms of "basic trust," as an individual disposition and a personal trait, as well as the long-term development of mutual trust relations. Psychological learning theories provide an answer to an important prerequisite of trust: the presence or absence of trust-related knowledge, and its generalization into schematic and typical knowledge structures, in the form of schemata and mental models. For that reason, learning theories are an indispensable ingredient in a broad theory of trust. Moreover, developmental models converge in their view that, at more mature stages of the trust relation, trust may become "blind," and move away from its cognitive basis. This indicates a noncognitive, irrational "leap of faith" in trust, which cannot be explained by sole reference to knowledge alone. Trust development is portrayed as occurring in qualitatively different stages, in which the prevalent type of trust changes from calculus-based to affect-based forms. Essentially, the typologies created in psychological developmental models point to an increasingly unconditional application of existing dyadic knowledge structures (i.e. relational schemata), paired with the rise of schema-triggered affect and the trustors' reliance on subjective emotional experiences as a "quick-step" to trust.

Sociological approaches, on the other hand, have emphasized the relational character of trust and the social embeddedness of the trust problem in a larger cultural and institutional context. For one, this means that trust-related knowledge is adopted during socialization by a contin-

ued internalization of the social stock of knowledge; it therefore is socially predefined. This opens up the avenue of regarding the prevalent "culture of trust" in a society, and the institutions which shape the way that trustors deal with the trust problem, by referring to the cultural and normative context as a source of trust-related knowledge. Institutions that help to establish trust come in the form of, for example, social norms, social roles, or habitual routines. Sociologists have advocated the view that action need not be instrumental on all occasions, and that the norms and rules, social roles and routines used to solve trust problems may override rational considerations if the actors follow them guided by a "logic of appropriateness." It is worth considering the functions that trust assumes in an individual sense and in a social sense. Individually, it can be understood as a mechanism for the reduction of social uncertainty; socially, it is a mechanism for the production of social capital and social integration. The critical functions that trust assumes in social interactions warrant that it often becomes institutionally protected—for example, by the establishment of norms of reciprocity and other moral norms (keeping promises, telling the truth, etc.). Thus, the institutional and cultural conditions that shape the emergence of trust must be respected when we think about how trustors solve trust problems. At the same time, sociological approaches emphasize an unconditional element in trust that is grounded on a different "logic" than the rational consideration of utility. It roots trust in the internalized institutional and cultural rules which actors can routinely apply, based on taken-for-grantedness, situational normality, and structural assurance.

Finally, the economic paradigm, which represents a current mainstream of trust research, demonstrates a vigorous attempt to formally specify and causally explain trust. In the rational choice approach, trust is essentially conceived of as a rational decision, made by trustors who engage in value-expectancy tradeoffs to discern the best alternative for proceeding, given the information, preferences, and constraints that they face. Modeling trust warrants that all parameters governing the choice of a trusting act must be formally captured and brought into a functional relation, so that the axiom of utility maximization can find its expression in a parsimonious and tractable model that yields equilibrium predictions. Embeddedness arguments can be easily recast in terms of additional cost and incentive parameters, and "wide" conceptions of rational choice incorporate psychological factors and social preferences to model the various motivations and encapsulated interests that a rational trustor might take into account. However, despite its formal clarity and precision, a huge body of empirical and theoretical evidence suggests that the rational choice paradigm is limited in its applicability to trust research. This lessens its attractiveness as a main explanatory vehicle. A critical factor is the question of rationality inherent in trust. While the rational choice paradigm is unambiguous in this regard, it clearly contrasts with the perspective of "blind" trust put forward by psychological and sociological researchers.

The chapter closes with a discussion and presentation of the main theoretical concerns and motivations of the present work: the relation between trust and rationality, which has been described inconsistently by different paradigms and research traditions. As will be argued, the neglect of the dimension of rationality in the trust concept is a main barrier to the theoretical integration of existing research. While the economic paradigm assumes the capability of actors to engage in considerations of a utility-maximization variety, sociological and psychological approaches emphasize that trust can be nonrational, in that actors apply the relevant knowledge and follow cultural and normative patterns automatically, based on taken-for-granted expectations and structural assurance. In a sense, these approaches portray the choice of a trusting act as being based on simple heuristic processes, substituting the ideal of rational choice with a "logic of appropriateness," in which the adaptive use of rules, roles, and routines helps to establish a shortcut to trust. The discussion of the relation between trust and rationality suggests that we have to turn to other theoretical paradigms which incorporate an individual actor's degree of rationality as a fundamental ingredient. Thus, we will have to answer the question of its mechanics and its links to the processes of interpretation and choice, while simultaneously specifying a clear and formally precise model.

3.1. Psychological Development

3.1.1. Learning and Socialization

A natural question to ask is whether trust can be "learned." In a trust problem, the trustor's expectation of trustworthiness depends on the information available to him; it depends on the way this stored knowledge is interpreted and used in conjunction with immediate situational impressions and affect. For example, knowledge of trustee characteristics depends on a concrete interaction history in which both actors have gotten to know each other and have had the opportunity to learn about each other's qualities. On the other hand, generalized expectations are synthesized from a multitude of experiences in the past, and help to inform immediate impressions of a potential trustee when specific knowledge is absent (often in the form of stereotypes). Learned social rules, norms, roles, and the like can inform the choice of a trusting act when their validity is indicated. In fact, all trust-related knowledge must eventually be learned—in short, "experience molds the psychology of trust" (Hardin 1993: 508).

In order to understand how knowledge evolves out of past experience, it is helpful to take a look at existing learning theories (see Anderson 1995). Broadly speaking, *learning* concerns the emergence of a connection, or association, between relevant features of a situation (stimuli) and the reactions of the organism (response). The two most important mechanisms for establishing such links are classical conditioning and instrumental learning. Classical conditioning concerns the association of stimuli and their internal evaluation in the form of experienced (dis)utility, which may be amplified by punishment and reward. Ultimately, this shapes the

preferences of an individual. Instrumental learning concerns the generation of causal hypotheses and theories about the structure of the world and about the effect of actions upon it. By "reinforcement" learning (the emergence, reinforcement, or extinction of existing associations) and "trial and error" learning, humans gradually pick up the contingencies of the environment, and use these to predict future contingencies and to plan actions. Ultimately, learned contingencies manifest in the form of expectations, which represent the mechanism through which past experiences and knowledge are connected to the future. With respect to expectations of trustworthiness, Hardin describes such an instrumental learning process as a "commonsense but likely unarticulated Bayesianism" (Hardin 1993: 508), in which past experiences are used to "update" expectations whenever new information can be added to the existing stock of knowledge.

In all learning, the processes of differentiation and generalization are important in determining the structure of knowledge and how it is organized. In essence, generalization and differentiation allow for the classification and typification of knowledge, that is, the discrimination of different domains and the abstraction of the "typical" from specific experiences. The resulting mental structures allow for the recognition of things which are already known, and which serve as interpretive schemes for reality. Any result of such an abstracting categorization of past experience, that is, typical knowledge and its mental representation, is called a *schema* (Rumelhart 1980).[1] Schemata facilitate the interpretation of events, objects, or situations and can emerge with respect to every aspect of subjective experience, concerning both our material reality and the world of thought. They vary in complexity, are often hierarchically organized, and can be conceived of as "organized representations of past behavior and experience that function as theories about reality to guide a person in construing new experience" (Baldwin 1992: 468). Importantly, as a building block of cognition, they are fundamental to the subjective definition of the situation. As Rumelhart puts it, "the total set of schemata instantiated at a particular moment in time constitutes our internal model of the situation we face at that moment in time" (1980: 37). Since the schema concept is very broad, researchers often devise more specific constructs according to their research program. For example, *identity* can be defined as a set of self-related schemata, including views about the self in relation to others (Greenwald & Pratkanis 1984). Likewise, a *stereotype* can be defined as a socially shared schema concerning the characteristic traits of a social category (Fiske 1993, Wheeler & Petty 2001). Terms that will be used synonymously for *schema* in the following are *mental model* and *interpretive scheme*.

[1] Rumelhart broadly defines a *schema* as "a data structure for representing generic concepts stored in memory. There are schemata representing our knowledge about all concepts: those underlying objects, situations, events, sequences of events, actions and sequences of actions. A schema contains, as part of its specification, the network of interrelations that is believed to normally hold among the constituents of the concept in question. A schema theory embodies a *prototype* of meaning. That is, inasmuch as a schema underlying a concept stored in memory corresponds to the meaning of that concept, meanings are encoded in terms of the typical or normal situations or events that instantiate that concept" (1980: 34).

In an attempt to find the optimal mix between differentiation and generalization, individual knowledge becomes structured along the dimensions of familiarity, clarity, determinacy, and credibility (Schütz 1967, Schütz & Luckmann 1973). While the boundaries of the individual life-world (circumscribed by spatial, temporal, and social distance) are always incomplete and potentially problematic, knowledge can also acquire a taken-for-granted character in which it is left unquestioned. This concerns those sectors of life in which actors frequently act, have detailed knowledge of, and where experience presents itself "as not in need of further analysis" (Schütz 1967: 74). The available schematic knowledge is then sufficient to solve the problems encountered in daily life, and there exist neither internal nor external motivations to further "update" or refine it. Even if this knowledge does not perfectly apply to a given situation, individuals can often bring back unfamiliar events into the familiar world ("familiarization") by recognizing their proximity to known schemata (Möllering 2006b).

Familiarity with the structures of the life-world lays the ground for *routine,* which develops out of regularly and habitually performed actions, and is rooted in "habitual knowledge." According to Schütz and Luckmann (1973), routine action is based on taken-for-grantedness, and it is directly related to the process of interpretation. A situation appears problematic and interrupts routine to the extent that the available knowledge is not sufficient to *define* it, that is, when "coincidence between the actual theme and the potentially relevant elements of knowledge does not occur sufficiently for the mastery of the situation in question" (Schütz & Luckmann 1973: 236). Normally, however, knowledge serves as a routine schema for action: "With routine coincidence, 'interpretation' is automatic. No explicitly judging explication occurs in which, on the one hand, the situation and, on the other hand, the relevant elements of knowledge come separately into the grasp of the consciousness to be compared to one another" (ibid. 198). This suggests that the application of trust-related knowledge also can become a matter of routine in familiar settings when taken-for-grantedness is in place.

In a broad perspective, learning, familiarization, and routinization are ever-present aspects of *socialization*, which describes the internalization of socially shared knowledge ("culture") and, coincidentally, the development of individual identity (Berger & Luckmann 1966). Culture presents itself to the individual as part of an objective reality, as an "inescapable" fact of life. During socialization, the basic rules of society, its obligatory norms and moral values, as well as the schematic knowledge of the "typical" and the "problematic," are internalized and learned from significant others (for example, parents). At the same time, the individual adopts and generalizes from experience a large set of socially shared interpretive schemes which can be used to attach meaning to typical situations (*frames*), typical actions (*norms, rules*), typical action sequences (*scripts*), and typical actions of typical actors (*roles*), along with a large amount of routine habitual knowledge (*routines*, or "knowledge of recipes") and technical skills. For example, a *role*, according to Berger and Luckmann (1966), is a socially shared

type of actor in a context where action itself is typified. Similarly, an *institution* is defined as "a reciprocal typification of habitualized actions by types of actors" (ibid. 54). All in all, this schematic cultural knowledge influences perception, interpretation, planning, and action (DiMaggio & Powell 1991, D´Andrade 1995). For example, frames, that is, the schematic knowledge of typical situations, help to focus the actors on the "primary goals" of an immediate situation (Lindenberg 1989, 1992). Most importantly, they provide the means for a reciprocal *social* definition of the situation. Through symbolic interaction, actors then negotiate and create a mutually shared meaning for the social situation (Mead 1967, Blumer 1969). This lays the ground for coordinated action and cooperation. Institutionalized cultural knowledge then becomes *externalized*, that is, "enacted" by schema application and the (potentially routine) execution of the actions prescribed by the relevant norms, roles, and scripts. Eventually, society's institutions are reproduced by its members on the basis of their everyday routine interactions.

In the discussion of both the objective structure and subjective experience of trust, a recurring theme was the knowledge that trustors can attend to when forming expectations of trustworthiness. The cognitive dimension of trust, and the trustor's knowledge of the social world, point to processes of learning, socialization, familiarization, generalization, and to the development of practically relevant interpretive schemes and their (routine) application. Many contributions suggest that there is a developmental path to trust and trustworthiness by which a learned "capacity to trust" (Hardin 1993) is forged. The ability to trust is based on past experience, learning, and familiarity with the individual life-world, which render available the different categories of trust-related knowledge: specific information, such as trustee characteristics, knowledge of dyadic and network embeddedness, and knowledge of the cultural-normative frameworks surrounding the trust relation—such as rules, roles, norms, and values, as well as generalized expectations, stereotypes, and so forth. The aspect of routinization in familiar and unproblematic sectors of life suggests that trust, building on relevant schemata, may become a matter of routine, too. Likewise, the fact that the knowledge acquired during socialization is by and large "socially conditioned" (Schütz & Luckmann 1973: 243f.) recasts the idea of institution-based and rule-based forms of trust. All in all, socialization and learning are important background processes that determine the antecedent conditions of interpersonal trust.

3.1.2. Basic Trust

A minimum of trust is necessary to live and to find one's bearing in life. However, this ability must be learned, and the essential qualifications for exhibiting trust have already been established in infants (Luhmann 1979: 27f.). Dyadic trust relations are prototypical because trust is first tested within the family and in the relations of infants to their attachment figures and significant others. In this sense, Erikson (1950, 1968, 1989) discusses the formation of *basic*

trust in infants and children. According to his psychoanalytic theory of personality development, each individual builds up a "basic sense of faith in the self and the world" (Erikson 1950: 80) within the first two years of life. The formation of this preconscious, diffuse sense of consistency and safety in the infant's relationship with a care-giver is, as Erikson points out, the first and most important task at the start of a human biography.

Basic trust develops with the child's experience that the environment provides for security and for the general satisfaction of needs; it is strongly influenced by the availability and intensity of parental care and the reliability and predictability of responses to the child. In terms of subjective experience, it is conceptualized as primarily emotional. It describes the innate understanding of being part of an ordered, meaningful social world and, a basic confidence in the future—ultimately, this lays the ground for the ability to see oneself connected to others through shared meaning, values, and norms (Lahno 2002: 325f.). As a part of the child's developing inner organization and identity, basic trust readily influences interactions with the social world. It evolves into a core orientation that others can or cannot be trusted, affecting the overall "readiness to trust" in interpersonal relationships. Notably, the success or failure of the development of basic trust has far-reaching consequences for infants in terms of emotional organization, self-perception, behavior, and coping capabilities in problematic or stressful situations (Scheuerer-Englisch & Zimmerman 1997).

In a similar fashion, Bowlby (1969) proposes that the earliest affective bonds formed by children with their caregivers are of primary importance and have a long-lasting impact that continues throughout life. His contributions have inspired a theoretical paradigm known as attachment theory (Bowlby 1969, 1973, Ainsworth et al. 1978, Bowlby 1980, Hazan & Shaver 1987, Ainsworth & Eichberg 1991, Cassidy & Shaver 1999). Attachment theory studies and explains the formation of attachment and relationship patterns over the life course. It revolves around the question of why and how infants become emotionally attached to their primary caregivers, and how and whether these early patterns of attachment transfer into individual relationship behavior and the development of relationships throughout the life course. The term *attachment* denotes an affective bond between an individual and an attachment figure, usually the parents or other caregivers. These bonds are based on the child's need for safety, security, and protection, and describe a "lasting psychological connectedness between human beings" (Bowlby 1969: 194).

Ainsworth et al. (1978) showed that early caregiving experiences translate into interindividual differences in the way children organize their attachment behavior. On an empirical basis, they identified three major attachment styles: secure, anxious-avoidant, and resistant. These *attachment styles* describe how infants relate to their environment in terms of approach-withdrawal behavior, how they cope with new and unexpected situations, and how they interact with others (Fraley & Spieker 2003). Children with the secure attachment style show min-

imal distress when left alone, use the attachment figure as a "secure base" to independently explore the environment, and respond with less fear and anxiety in novel situations. In contrast, children without the secure attachment style display more fearful, angry, and upset behaviors than do the securely attached children, are less independent, and more frequently use withdrawal strategies to cope with problematic situations.

According to attachment theorists, the experience of regularity, attentiveness, responsiveness, tactfulness, and empathy in parental care leads to secure attachment—an innate sense of assurance and confidence which is similar to Erikson's concept of basic trust. Infant attachment behavior is guided by an internal attachment system which follows the general goal of maintaining "felt security" (Bretherton 1985). Beginning with the first months of life, the infant's experiences with the caregiver are memorized, and lay the ground for the development of "internal working models" about social relationships and the environment.[2] Internal working models are schemata of the self, of others, and of the environment, which help individuals to predict and interpret situations and behavior (Pietromonaco & Barrett 2000). Sticking to our terminology, we will use the term *relational schema* to denote such a cognitive structure (see Baldwin 1992, Chen et al. 2006). The feeling of security and faith in the reliability and responsiveness of parental care which characterizes basic trust is a by-product of one of the first relational schemata developed by infants (although it is not as differentiated and specific as the relational schemata developed in the later course of life). Basic trust and secure attachment mirror the relationship experience of the infant; they reflect an inner security which comes with the ongoing confirmation that "everything is normal" and is continuously "as expected." Therefore, we can interpret basic trust as the first emotional experience of the familiarity and "taken-for-grantedness" of reality in the emergence of the natural attitude to the life-world.

The findings of developmental psychology concerning the formation of basic trust in early childhood are important, because they point towards the origins and antecedent conditions of interpersonal trust. Bowlby hypothesized that attachment behavior characterizes human beings "from the cradle to the grave" (1979: 129). He suggested that the relational schemata and attachment styles which are formed early in life generalize and extend to adulthood, and shape interpersonal attachment behavior with significant others and peers, and in close relationships. In fact, an increasing amount of empirical research indicates that the links between basic trust and early and adult attachment styles do exist, although the precise mechanisms which explain their stability and their changes over time are not fully understood (Fraley 2002, 2010).

[2] "Each individual builds *working models* of the world and of himself in it, with the aid of which he *perceives* events, *forecasts* the future, and constructs his plans. In the working models of the world that anyone builds a key feature is his notion of who his attachment figures are, where they may be found, and how they may be *expected* to respond. Similarly, in the working model of the self that anyone builds a key feature is his notion of how acceptable or unacceptable he himself is in the eyes of his attachment figures" (Bowlby 1973: 203, emphasis added). As Baldwin (1992) points out, the working model concept is identical to the concept of a *relational schema*.

For example, Hazan and Shaver (1987) conceptualize romantic love in close relationships as an attachment process similar to infant attachment. In their study, they empirically identify the same patterns of attachment styles in adults which Ainsworth et al. (1978) had found in children. Secure adult attachment styles tend to be associated with relationships that are characterized by higher levels of interdependence, trust, commitment, and satisfaction (Simpson 1990, Shaver & Hazan 1993). Importantly, adult attachment styles are tightly connected to differential relational schemata of the self and others; they involve different views of romantic love and love-worthiness, as well as different expectations about the availability and trustworthiness of partners. While secure attachment types tend to seek closeness and intimacy, and are open to new relations with others, the anxious and avoidant types experience insecurity about other's intentions, fear both intimacy and being unloved, and prefer distance and independence.

Secure attachment styles are also positively related to measures of generalized interpersonal trust (Collins & Read 1990). Mikulincer (1998) finds that adult attachment styles are directly connected to the subjective experience of trust. Attachment style groups differ in the level of trust they feel towards partners, in the accessibility and affective quality of trust-related memories, in the appraisal of trust-related experiences, relationship goals, and in the strategies of coping with a breach of trust. In sum, attachment styles can be viewed as referring to "differences in the mental representations of the self in relation to others, to particular types of internal working models of relationships, models that direct not only feelings and behavior but also attention, memory and cognition" (Main et al. 1985: 67). This suggests that the development of early basic trust and the corresponding relational schemata may have long-lasting effects and accumulate into "a more stable trust orientation that may be activated and applied in close relationships" (Mikulincer 1998: 1221).

3.1.3. Individual Dispositions and Traits

The idea that the learning of trust-related knowledge can evolve into a stable disposition, or personality trait, has a long standing in trust research (Erikson 1968, Rotter 1980, Hardin 1993, 2002). Theories of "dispositional trust" (Kramer 1999), focusing on interindividual differences in trusting behavior, have been proposed regularly in the area of psychological trust research. Central to dispositional theories of trust is the assumption that certain factors within individuals predispose them to trust or distrust others.[3] A prominent example of such an account is the work of Rotter (1967, 1971, 1980), who regards trust as a generalized expectation of the trustworthiness of others, and develops an attitudinal measure to measure its impact (the

[3] Authors have synonymously used the terms *trust propensity* (Mayer et al. 1995), *disposition to trust* (McKnight et al. 1998), *general trust* (Yamagishi & Yamagishi 1994), *global trust* (Couch & Jones 1997), and *faith in humanity* (Wrightsman 1974, 1991) to generally denote "the extent to which one believes that non-specific others are trustworthy" (McKnight et al. 1998: 478).

"Interpersonal Trust Scale," ITS). Arguing in the context of social learning theory, he posits that individuals develop a generalized expectation of other people's trustworthiness, or *generalized trust*, in response to their personal history of trust-related experiences over their life-course. This is achieved by generalization, differentiation, and reinforcement learning. On top of that, individuals adopt relevant cultural schemata from significant others or from the mass media.[4] Generalized trust can be regarded as the default expectation of the trustworthiness of unfamiliar others, which influences how much trust one has for a trustee in the lack of any other specific information. The influence of such a generalized expectation increases with the novelty, unfamiliarity, and atypicality of the situation; *vice versa*, specific expectations can replace generalized expectations and determine the choice of a trusting act, once trustor and trustee become acquainted with each other (Rotter 1980, Johnson-George & Swap 1982).[5]

Hardin (1993, 2002) develops a similar argument, stating that individuals come to know about the general trustworthiness of others using naïve Bayesian learning. As a result, each individual develops an idiosyncratic "capacity to trust." On the macro-level, different types of trustors emerge. Recasting the arguments of attachment theory, Hardin maintains that experiences in the early years of life (e.g. neglect, abuse, trauma) highly influence the individual development of the capacity to trust, which represents "general optimism about the trustworthiness of others" (1993: 508). Low-trust types are at a double disadvantage: they cannot capitalize from trust directly in terms of utility (that is, they risk neither gain nor loss), and they neglect the learning and updating opportunities to test whether their overly negative expectations of others are justified. On the other hand, high-trust types will enter interactions more frequently, but may suffer severe losses if they are too optimistic. However, Bayesian updating suggests that these types can quickly readjust their expectations to an optimal level that aptly reflects the conditions of the social environment, whereas low-trust types suffer from an ever-increasing relative disadvantage.

In this line, a number of empirical studies report differences between high and low-trust individuals. Generally, high-trust types tend to be more sensitive to trust-related information and more accurate in judging the trustworthiness of others (Yamagishi et al. 1999). As suggested by Hardin, they also adjust more quickly to the threat of a breach of trust and signs of untrustworthiness (Yamagishi 2001). High and low-trust individuals differ with respect to the attribution of motives to a trustee and to the interpretation of responses (Holmes 1991, Jones

[4] Thus, Rotter relates the development of a disposition to trust not only to learning from personal experience, but also to the acquisition of cultural schematic knowledge (i.e. stereotypes) from significant others and from mass media. Notably, the idea of an intergenerational transmission of trust-related attitudes has recently received empirical support (Dohmen et al. 2006).

[5] For example, Johnson-George and Swap conclude that "disposition to trust" predicts the choice of a trusting act *only* in "highly ambiguous, novel, or unstructured situations, where one's generalized expectancy is all one can rely on" (Johnson-George and Swap 1982: 1307).

& George 1998, Rempel et al. 2001). High-trust individuals have been found to behave more trustworthily and honestly (Rotter 1971, 1980), perceive interpersonal relations as less problematic and distressful (Gurtman 1992), and are often regarded favorably by others (Rotter 1980). Nevertheless, high levels of generalized trust do not equate to *gullibility*, that is, to a naïve and credulous belief which "overestimates the benignity of other people's intentions beyond the level warranted by prudent assessment of available information" (Yamagishi & Yamagishi 1994: 135). On the contrary, high generalized trust can be regarded as a result of an individual's cognitive investment into detecting signs of trustworthiness in environments with social uncertainty and risk—and as a consequence, the skills needed for discerning trustworthiness develop and become more refined. In short, high levels of generalized trust may be indicative of an improved social intelligence (Yamagishi 2001).

The empirical evidence concerning the direct relation between generalized dispositions to trust and overt trusting behavior is, however, mixed. In an extensive meta-analytic study, Colquitt et al. (2007) show that dispositions to trust have an influence on trust-related outcomes, such as risk taking, task performance, and organizational citizenship behavior, but these relations are only moderate. At the same time, it has been found that single-item measures of generalized trust do not successfully predict trusting behavior in experiments (Glaeser et al. 2000, Naef & Schupp 2009) or in close relationship contexts (Larzelere & Huston 1980). Holmes (1991) posits that the link between "generalized tendencies" toward trust and its development in particular relationships has not been directly established. This follows from the empirical observation that trust is often highly dependent on the situational context. Extending the model of Mayer et al. (1995), McKnight et al. (1998) accommodate for this fact by treating "disposition to trust" as only one antecedent factor among others to influence expectations of trustworthiness and the willingness to take risks. Their model attempts to explain the regularly high levels of initial trust between strangers, which, according to the authors, present a "paradox." Notably, they argue that dispositional tendencies can have a direct effect on trusting intentions, but may be mediated (and outweighed) by institution-based forms of trust and cognitive processes such as stereotyping and categorization—including reputational effects. In line with Rotter (1980) and Johnson-George and Swap (1982), they argue that dispositional tendencies to trust will have an effect primarily in new relationships or in one-shot situations with strangers—that is, when more specific situational information is not available.

Empirically, Gill et al. (2005) show that individual dispositions to trust, as measured by a modified version of Rotter's ITS, predict trusting intentions only "when information about the trustee's ability, benevolence, and integrity is *ambiguous*" (ibid. 292, emphasis added). Building on the work of Mischel (1977), the authors introduce *situational strength* as a boundary

condition for the relation between dispositional trust and trusting intentions.[6] If a situation contains strong cues about the trustworthiness of a potential trustee (cues can relate to many sources of trust-related knowledge, ranging from individual characteristics to institutional and normative structures), then individual dispositions and traits step back in favor of the available evidence and more specific trust-related knowledge.

All in all, the development of a stable disposition to trust, much like a "personality trait," seems to be one important factor in determining the build-up of trust in a particular trust problem. Dispositional differences between individuals have a measurable effect on a variety of trust-related constructs, and therefore must be respected when explaining the emergence of trust in a particular trust relation. However, it is important to keep in mind that external factors ("situational strength") may moderate the impact of such generalized dispositions. Their influence in a particular context may be limited, and is itself highly context-dependent.

3.1.4. Models of Trust Development

An important stream of trust research focuses on the development and change of interpersonal trust in ongoing relationships (Lewis & Weigert 1985b, Rempel et al. 1985, Holmes 1991, Lewicki & Bunker 1995b, McAllister 1995, Jones & George 1998, Lewicki et al. 2006, Ferrin et al. 2008). Central to these models is the assumption that relationships continue over an extended period of time. By repeated interaction and iterated exchange, actors can develop multiple trust relations with each other, and acquire very specific trust-related knowledge. This research focuses on the evolution of expectations, intentions, and affect towards the other over time, as well as the perceptions and attributions of trustee characteristics, moral qualities, and motives. Furthermore, developmental models often implicitly assume a switching of roles between trustor and trustee, so that trust relations become reciprocal. In such a dynamic setting, the growth and decline of interpersonal trust within ongoing relationships is analyzed.

To characterize the status of relationships, Lewicki et al. (1998) introduce the terms "relationship bandwidth" and "relationship richness." Relationship bandwidth describes "the scope of the domains of interpersonal relating and competency that are relevant to a single interpersonal relationship [....] The broader the experience across multiple contexts, the broader the bandwidth" (ibid. 442). In the extreme case, a relationship with a very narrow bandwidth might offer only one opportunity to maintain a trust relation of the form "A trusts B with re-

[6] "According to Mischel (1977), situations can be characterized on a continuum from *strong* to *weak*. Strong situations have salient behavioral cues that lead everyone to interpret the circumstances similarly, and induce uniform expectations regarding the appropriate response. [...] Thus, strong situations are said to suppress the expression of individual differences. Weak situations, on the other hand, have highly ambiguous behavioral cues that provide few constraints on behavior, and do not induce uniform expectations. [...] In weak situations, the person has considerable discretion in how to respond to the circumstances. Thus, weak situations provide the opportunity for individual differences such as personality to play a greater role in determining behavior" (Gill et al. 2005: 293).

spect to X." On the other hand, a large bandwidth permits the emergence of multiple trust relations, so that "A trusts B with respect to X, Y and Z." Moreover, relationship *richness* describes the "texturing of relationships" (ibid.), that is, the details of knowledge across the bandwidth. Relationship richness increases in an ongoing relationship because the parties acquire more information about each other. This warrants trust becoming "fine-grained" (Gabarro 1978) and differentiated with respect to each unique trust relation.

Most dynamic models assume that trust starts at a low level and builds up incrementally over time as a result of experience (Lewicki et al. 2006). Generally speaking, the propellants behind this gradual increase in trust are "mutually satisfying interactions" (Rempel et al. 1985). Changes in the level of trust are driven by the experience of rewarding or punishing outcomes, defined by the incentive structure of the trust problems encountered during repeated interaction. These outcomes shape the subjective experience of trust in the form of individual cognition and affect: "Successful behavioral exchanges are accompanied by positive moods and emotions, which help to cement the experience of trust and set the scene for the continuing exchange and building of greater trust" (Jones & George 1998: 536). Importantly, a symbolic communication of trustworthiness perceptions alone is not sufficient to create an upward spiral of mutually reinforcing levels of trust; the actual taking of risks (in the choice of trusting acts) and the observable cooperative responses (in the honoring of the trust) are necessary to fuel this development (Zand 1972, Ferrin et al. 2008). At the same time, "successful" interactions open up the opportunity for further engagements, and warrant increases in the bandwidth and richness of the relationship—in this sense, "trust breeds trust," as the actors can test the validity of their initial trustworthiness judgments and correspondingly increase mutual vulnerability and dependence when trust is honored.

Models differ as to whether the evolution of trust is regarded as a continuous process or as a succession of discrete developmental stages. Continuous accounts, regularly proposed in the cognitive tradition of trust research, conceive of development as a Bayesian learning process in which the actors gradually accumulate trust-related knowledge and withdraw trust when it is failed (Deutsch 1958, Hardin 1993, Kramer 1996). Thus, specific expectations evolve out of past experience, and actors adjust their level of trust by updating these expectations based on the available evidence of trustworthiness. The emergence of specific expectations can be recast as an augmentation of relationship richness, so that more precise judgments can be made within each unique trust relation. In other words, the ambiguity of expectations decreases over time. Naturally, when favorable expectations of trustworthiness have developed, this opens up an avenue to increase relationship bandwidth and—more generally—to extend the scope of cooperation (Zand 1972, Ostrom 1998)

In contrast, a recurring theme in many psychological models of trust development is that interpersonal relationships can be characterized by qualitatively distinct stages, and that the na-

ture of trust, as well as its basis and subjective experience, change as the relationship matures (Rempel et al. 1985, Jones & George 1998, Lewicki et al. 2006). These stages are often regarded as hierarchical: in terms of time and emotional engagement, each stage requires additional investments by the actors involved (Rempel et al. 1985). For example, Lewicki and Bunker (1995) envision the evolution of trust as a sequence of hierarchical stages which they denote as (1) calculus-based trust, (2) knowledge-based trust, and (3) identification-based trust.

(1) In the first stage, calculus-based trust, the choice of a trusting act is accompanied by a calculation of the potential costs and benefits which the trustee (!) would incur by choosing to be untrustworthy. This amounts to a rational consideration of what Hardin (2001, 2002) describes as "encapsulated interest." Essentially, the choice of a trusting act is justified in the trustee's incentive to be trustworthy, and this incentive is grounded in the future value of maintaining the relationship. This also includes considerations of costs and benefits "outside" of the particular relationship—for example, a potential loss of reputation. In a similar fashion, Rempel et al. (1985) state that the first stage of trust development in intimate relationships ("predictability") resembles a forecast of the partner's future actions, which relies on an understanding of the "reward contingencies underlying potential actions" (ibid. 97).

(2) By repeated interaction, expectations stabilize and relationships increase in bandwidth and richness, augmenting the value of the relationship itself. A shift to the second stage of knowledge-based trust ("dependability," Rempel et al. 1985) occurs once the actors become solidly acquainted to each other and extend the relationship scope. This means that the actors acquire precise estimates of their characteristics, moral qualities and underlying value systems. As Jones and George (1998) argue, ongoing interactions and positive affective experiences accompanying successful cooperation are also conducive to the emergence of shared interpretive schemes and the development of a common "frame of reference." According to Lewicki et al. (2006), the interaction frequency, duration, intensity, and diversity of the challenges which the actors overcome in the ongoing relationship determine the point in time at which calculus-based trust shifts to knowledge based trust. Importantly, this stage includes an increased level of "attributional abstraction" (Rempel et al. 1985)—the focus of the trustor's expectations moves away from assessments of direct consequences of specific actions to an overall evaluation of the qualities and characteristics of the trustee.

Jones and George (1998) subsume these two stages under the rubric of *conditional trust*, a type of trust which is "consistent with the idea that one of the bases for trust is knowledge" (ibid. 536). Notably, conditional trust includes only a "pretense of suspension." Although a trustee chooses a trusting act, this does not mean that uncertainty is internally removed at this stage, even though the trustee acts as if this were the case. Rather, the choice of a trusting act is often simply preferable to initial distrust, because it saves cognitive resources and allows

for a "tit-for-tat" retribution strategy at the same time (Deutsch 1958, Luhmann 1979). As the name suggests, conditional trust is prone to being withdrawn when expectations are failed, and quickly updated in a case of failure ("trust but verify"). According to Jones and George (1998), conditional trust is sufficient to enable most social interactions.

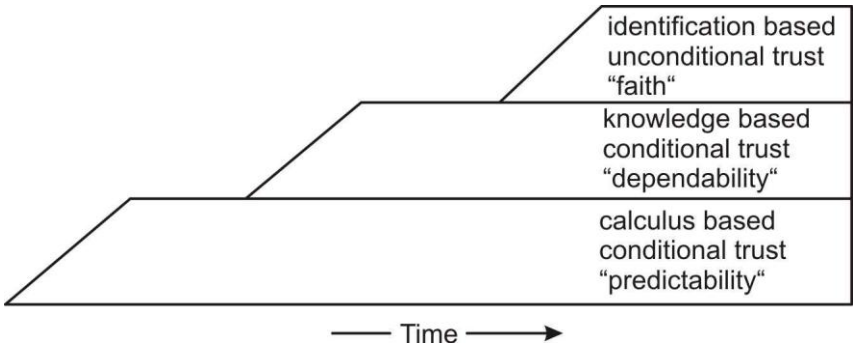
(3) The third and last stage of "identification-based" trust occurs when the parties develop mutual identification with, and strong affect towards, each other. On the one hand, this means that the available knowledge is sufficient to induce a "full internalization of the other's preferences" (Lewicki et al. 2006: 1009), which facilitates the development of common goals and shared values, and brings about a motivational change from pursuing self-interest towards maximizing joint outcomes. On the other hand, mutually developed and shared interpretive schemes now structure the subjective definition of the situation to such an extent that trustworthiness is regarded as unquestioned, "based on confidence in the other's values" (Jones & George 1998: 537). This stage of *unconditional trust* signifies a "real" suspension of uncertainty, which replaces the mere "pretense" thereof in conditional trust, and goes along with an increase in mutual attachment. Importantly, the attributions made to the other's motivation now emphasize intrinsic (as opposed to external or instrumental) motives, such as the shared enjoyment of activities, the demonstration of affect, a sense of closeness, and a shared social identity that is established (Rempel et al. 1985). This third and last stage ("faith," ibid.) exemplifies the real "leap of faith" that has occurred. The actors fully suspend uncertainty and doubt, and it is with a subjective certainty and emotional security that the relationship continues into the future—coincidentally, trustworthiness now is "taken for granted."

It is apparent that these developmental accounts closely resemble the distinction between cognition-based and affect-based forms of trust introduced earlier. Rousseau et al. (1998) differentiate "calculus-based trust" from "relational trust" and directly relate these types to McAllister's (1995) distinction of cognition and affect-based trust (see chapter 2.2.4). While calculus-based trust is assumed to rely on rational decision-making processes and on the principle of "trust but verify," relational trust derives from repeated interactions between trustor and trustee, which fosters the development of concern and emotional attachment. Likewise, Kramer (1999) holds that approaches to trust follow either a "rational choice" or a "relational" perspective, arguing that trust needs to be conceptualized "not only as a calculative orientation toward risk, but also as a social orientation toward other people and toward society as a whole" (ibid. 573), including the consideration of "self-presentational concerns and identity-related needs" (ibid. 574), which may influence the subjective experience of trust and subsequently the choice of a trusting act.

An important aspect in most developmental models is the fact that the perspectives of the actors involved change over time. Lewicki and Bunker (1996) describe the shifts between the different stages as "frame changes," that is, changes in the prevalent means of interpersonal

perception.[7] Jones and George (1998) highlight the importance of the development of shared interpretive schemes during repeated interaction. According to these authors, trust is experienced through changing attitudes, which they define as "(1) the knowledge structures containing specific thoughts and feelings people have about other people, and (2) the means through which they define and structure their interactions with others" (ibid. 533). We have earlier introduced the concept of a relational schema to denote such a cognitive structure.[8] Generally speaking, the succession of different stages in trust development can be recast as the creation, modification, and gradual enrichment of specific relational schemata, serving as a source of trust-related knowledge and "framing" a shared definition of the situation. Like any other type of knowledge, these knowledge structures can acquire a taken-for-granted character and become a matter of routine in familiar settings (figure 7).

Figure 7: Stages of trust development, adapted from Lewicki & Bunker (1996: 156)



Overall, models of trust development exemplify and extend our understanding of the antecedents of interpersonal trust by highlighting the development of an important source of trust-related knowledge: specific relational schemata, corresponding favorable expectations, and their associated affect and attachment. At the same time, developmental models highlight the fact that trust is related to different "modes" of subjective perception, ranging from more calculative orientations to a securely rooted state of affect paired with suspension and mutual identification—they are indicative of a flexible degree of rationality involved in a trusting act. As Hardin (1993) correctly points out, it is important to keep in mind that "thick-relationship theories" of trust merely display one possible source of trust-related knowledge, and one

[7] "The shift from calculus-based trust to knowledge-based trust signals a change from an emphasis on differences or contrasts between self and other (being sensitive to risk and possible trust violations) to an emphasis on commonalities between self and others (assimilation). The shift from knowledge-based trust to identification-based trust is one from simply learning about the other to a balance between strengthening common identities while maintaining one's own distinctive identity in the relationship" (Lewicki et al. 2006: 1012).

[8] In psychological research, the concept of *attitude* is used much more generally, including as it does affective, cognitive, and behavioral orientations towards *any object*, whether material, immaterial, person, or "thing" (see Olson & Zanna 1993). Thus, relational schemata are more specific constructs than attitudes, restricted to interindividual orientations in a social context. They represent a combination of schemata towards the self, the other person, and the relationship in question, as well as *interpersonal scripts*, including expectations of thoughts, feelings, goals, and actions of both the self and the other (Baldwin 1992).

source of incentives for the trustee to be trustworthy. This does not give them conceptual or theoretical priority over other sources and related theoretical accounts. A general theory of trust must, however, contain "trust in thick relationships" as a special case.

3.2. Sociological Perspectives

3.2.1. Functions of Trust

Sociological explanations of trust do not focus on the individual learning processes by which trust-related knowledge is acquired and accumulated into dispositional tendencies. They ask instead for the role that trust plays in the context of a human reality which is fundamentally social. This question naturally relates trust back to the social environment in which it is embedded. Theories of learning and development create a necessary "input" from which this analysis can be carried out, whereas sociological conceptions stress the relational character of trust, in the sense that trust "must be conceived as a property of collective units (ongoing dyads, groups, collectivities), not of isolated individuals" (Lewis & Weigert 1985b: 968). Trust, when mutually structuring subjective experience and action, is a property of the social system under scrutiny; in short, "the cognitive content of trust is a collective cognitive reality that transcends the realm of individual psychology" (ibid. 970). Its function is primarily sociological because it is not needed outside of social relations. Accordingly, trust is regarded as an elementary precondition for a wide range of social processes. It presents a core phenomenon for sociological thought and theorizing (e.g. Garfinkel 1963, Blau 1964, Luhmann 1979, Durkheim 1984, Lewis & Weigert 1985a, Coleman 1990, Giddens 1990, Endress 2002, Möllering 2006b).

In order to understand the relevance of trust for the functioning of social systems at large, it is necessary to inspect the role which trust plays in social processes. As noted previously, the concepts of trust and familiarity point to the experience of a taken-for-granted life-world and indicate the acceptance of a large part thereof as an implicit background assumption for further action. The routine and implicitness of social life occurs, however, in face of the ever-present possibility of a breakdown of social reality as it is known, a crumbling of taken-for-grantedness and of the routine "frames of reference" (Garfinkel 1963). Trust and familiarity constitute one important interactional resource preventing such a breakdown. In the words of Luhmann (1979), the need for trust emerges in face of the "complexity" of the social world through which each individual must navigate. This complexity needs to be resolved in order for individuals to remain capable of acting. Luhmann argues that trust is the most important *psychological mechanism for the reduction of social complexity*. We have already specified this social complexity when elaborating on the objective structure of trust: every basic trust problem includes irreducible social uncertainty. The interdependence of humans with respect to actions and outcomes (the "double contingency") in most social situations (and in trust

problems in particular) constitutes a source of social uncertainty, and, incidentally, the source of social complexity.

To resolve social complexity means to reduce the set of possible actions in face of contingent consequences, and to plan a course of action. As a psychological mechanism for the reduction of social complexity, trust "goes beyond the information it receives and risks defining the future" (ibid. 20). Trust is in place when favorable expectations initiate the choice of a trusting act, and it bridges existing uncertainty by fixing a definite future as a viable option. The formation and stabilization of expectations are therefore central processes in the reduction of complexity; likewise, they are central functions of trust. Notably, Luhmann argues that the learning, generalization, and development of mental schemata which abstract from reality are crucial elements that allow for such a functional reduction of complexity. While abstracted representations of the outside world work at a lower level of complexity than the actual environment, this implies at the same time that they "exhibit fewer possibilities, or more order" (ibid. 26) than the environment. Luhmann describes the reduction of complexity as a change in the level at which uncertainty is made tolerable—with trust, external uncertainty is substituted by inner certainty in a movement towards "indifference."[9]

At the same time, the suspension of social uncertainty results in a truncation of further searching and retrieval processes and reduces individual cognitive load. Therefore, trust is regularly conceived of as an efficient strategy to deal with scarce cognitive resources (Lewis & Weigert 1985b, Ripperger 1998: 258). By "extrapolating" past experiences into the future individuals save the cognitive resources which would be otherwise needed for the search of information and its deliberate processing. Consequentially, Lewis and Weigert (1985b) hold that trust is an alternative to rational prediction, reducing complexity "far more quickly, economically, and thoroughly." This is because rational prediction, in face of high social uncertainty, is costly, time-consuming, in principal limitless, and may "complicate" decision making. On top of that, "information may reduce, but cannot entirely eliminate, perception of uncertainty about future results" (Lewis & Weigert 1985a: 462). Their proposed answer is that trust allows actors to act "as if*"* certain futures are not possible (*viz.* suspension). However, the reduction of complexity and social uncertainty by suspension in a given situation necessitates that other circumstances are regarded, *ceteris paribus*, as unproblematic. Familiarity, as has been argued, is a precondition to trust. It is the power of trust and familiarity to effectively reduce social complexity that qualifies them as sociological core phenomena and as a basis for almost all social

[9] "Trust, by the reduction of complexity, discloses possibilities for action which would have remained improbable and unattractive without trust. For this reason, the benefit and rationale for action on the basis of trust are to be found ... in, and above all, a movement towards *indifference*: by introducing trust, certain possibilities of development can be excluded from consideration. Certain dangers which cannot be removed but which should not disrupt action are neutralized" (Luhmann 1979: 25).

processes. In their absence, social action would be paralyzed by the intrusion of the enormous complexity of a contingent social world and by the unpredictability of the future.

But sociological approaches to trust do not only scrutinize its functionality with respect to the individual capability of action. A second prominent theme is the analysis of the functions that trust performs with respect to the social systems in which it is developed and sustained. Many scholars regard trust as an indispensable ingredient for the functioning of social systems in general (Lewis & Weigert 1985a, b, Misztal 1996). Trust is seen as an efficient mechanism governing both market and nonmarket transactions (Arrow 1974, Bromiley & Cummings 1995), and a sort of "ever-ready lubricant that permits voluntary participation in production and exchange" (Dasgupta 1988: 49). According to Sztompka, it "encourages sociability, participation with others in various forms of associations, and in this way enriches the network of interpersonal ties" (1999: 105). By favoring communication, it also "encourages tolerance, acceptance of strangers, recognition of cultural or political differences as legitimate ... bridles expressions of inter-group hostility and xenophobia, and civilizes disputes" (ibid.). In this line, Ripperger (1998) concludes that one primary systemic function of trust is the generation of social capital. It constitutes an "organizing principle" (McEvily et al. 2003), in that it structures interaction patterns, stabilizes social structure, and mobilizes actors to contribute, combine, and coordinate resources toward collective endeavors. On an even more fundamental level, it is regarded as a prime ingredient in the successful social integration of modern society at large and the maintenance of social order (Luhmann 1988, Giddens 1990, Misztal 1996).

In asking for the social and systemic functions of trust, and in trying to understand the role that trust plays in social systems, it is necessary to extend the scope of the trust relation beyond the narrow frame that has been adopted so far. Trust, as a property of social systems, has to be understood in a much richer setting than that of a unidirectional trust relation devoid of context. By introducing the actor's social embeddedness (chapter 3.2.2) and discerning the connection between trust and social capital (chapter 3.2.3), we will advance our understanding of the functions of trust for social systems at large (chapter 3.2.4) in the following sections.

3.2.2. Social Embeddedness

If trust is primarily a social phenomenon, then it cannot be understood without reference to the social structures surrounding the trust relation. As Luhmann argues, "trust occurs within a framework of interaction which is influenced by both personality and social system, and cannot by exclusively associated with either" (1979: 6). Therefore, an adequate theory of trust must bridge micro, meso, and macrolevels of analysis. Purely cognitive models of trust development provide a "necessary but not sufficient understanding of trust phenomena" (Kramer 1999: 572), because trust emerges in a world that is rich in cultural meaning. Essentially, trust relations are not simply dyadic phenomena between two actors, but they normally occur with-

in a larger context, often possess a history, and may be influenced by other actors and institutions. Social embeddedness influences the strategies which trustors will use to solve a trust problem because it affects the availability of resources, determines the direct and indirect costs of action (that is, it influences the incentive structure of the trust problem) and governs the activation of norms and other cultural schemata (Heimer 2001). Social embeddedness includes relationships between a trustor and trustee (such as repetition, an interaction history, or the distribution of power), between the trusting parties and other members of a social system (for example social networks, reputation, group membership), and between actors and the relevant social system or its properties (normative structure, cultural practices, a "climate" of suspicion or trust, legal frameworks etc.). Consequentially, when asking for the sources of trust-related knowledge which a trustor can use, and when thinking about how this knowledge will be used, we have to take into account the social embeddedness of trustor and trustee.

Broadly speaking, *social embeddedness* refers to the constraining effects of ongoing social relationships on individual action (see Granovetter 1985, Uzzi 1997, Buskens & Raub 2008). The concept of embeddedness is based on the idea that actors must not be regarded as "atomistic" decision-makers, but as being embedded in networks of personal relationships—action always takes place in a social context. Networks of relationships between actors exert an influence on trust and trustworthiness primarily through the mechanisms of *learning* and *control* (Yamagishi & Yamagishi 1994, Buskens & Raub 2002). Control implies that direct or indirect reward and punishment opportunities are available in response to the actions of the trustee. With control, the incentive structure of the trust problem changes, such that the long-term value of trustworthiness is higher to the trustee than the short-term gains of failing trust. Learning, on the other hand, is the mechanism by which a trustor can acquire more information about the trustee, either directly by past interaction, indirectly from third parties via reputation, or via the surrounding institutional structures—for example, from social roles or norms. As with micro, meso, and macrolevels of analysis, social embeddedness is differentiated into dyadic, network, and institutional embeddedness (see Buskens & Raub 2008).

Dyadic Embeddedness refers to repeated interactions between two actors. It points to the temporal-structural aspect of embeddedness, denoting a situation in which a history of interactions between the trustor and trustee already exists, or in which trustor and trustee will likely face each other again in the future. To begin with, repeated interaction, if the "shadow of the future" is high enough, may increase the value of an ongoing relationship for the parties involved, and persuade even purely self-interested actors of the advantages of conditional cooperation, because the long-term benefits of continuing the relationship outweigh the short-term incentives for defection (Trivers 1971, Axelrod 1984). At the same time, the trustor can exert influence over the trustee because a failure of trust can be sanctioned by a withdrawal of future trust ("dyadic control"). If the incentive for abusing trust is not too high, this can give rise

to an equilibrium in which trust is always placed and always honored (Kreps 1990). A common term used is to describe such conditional cooperation is *weak reciprocity* (Gintis 2000c, Fehr & Schmidt 2006, Fehr & Gintis 2007). Weak reciprocity can be motivated by the long-term, "enlightened" self-interest of the players; essentially, it requires that reciprocal strategies are profitable and maximize the players' payoff in the long-rung.[10] This does not mean, however, that the trustor can enforce trustworthiness or that behavior becomes deterministic. Although expectations may become favorable and confident with dyadic embeddedness, a trust problem structurally requires a transfer of control over resources or events specified by the content of the trust relation—that is, even with dyadic control, a trustee might principally fail trust.

The argument points, however, to the interesting relationship between trust and power. If the trustor did possess a large incentive to continue the trust relationship, then the threat of abandoning the relationship would not be credible. As Farrell (2004) notes, trust relationships can endure a certain amount of asymmetry in the distribution of power without leading inevitably to distrust.[11] On one hand, ongoing relationships— especially when the actors have already invested many resources and developed emotional attachment ("sunk costs")—are less likely to be highly asymmetrical in power, and will be of value for both parties. On the other hand, when power asymmetries exist, then the actor who has less interest in the continuation of the relationship has more power, in the sense that his threat of exit is more credible; consequentially, he has less "need" to be trustworthy. This can hamper the development of trust, because (a) less powerful actors will often misconstrue and misinterpret the intentions of the more powerful one ("paranoid cognition," Kramer 2004), (b) actors may have different time horizons, in that the more powerful actor's time horizon is shorter and more limited, and (c) asymmetries of power make it more difficult to coordinate a mutually beneficial equilibrium, because the more powerful actor has the incentive to renegotiate over the outcomes of cooperation (Farrell 2004). These circumstances increase the social uncertainty that the less powerful actor has to face, and may lead to a point where trust is not possible anymore. When one actor is much more powerful than the other, he has no incentive to take into account the other's interests, has no reason to be trustworthy, and he is incapable of making credible commitments. Similarly, the less powerful actor has no incentive to be trustworthy, knowing that the other

[10] In contrast, *strong reciprocity* describes intrinsically motivated behavior (costly punishment, or cooperation even when defection would maximize payoffs) based on other-regarding preferences, which appears suboptimal to standard game theory, but can be accommodated for in psychological game-theoretic models (see chapter 3.3.4.).

[11] Farrell defines *power* in the context of bargaining situations: "Parties who have many possible attractive alternatives should a particular relationship not work out will be more powerful than parties who have few such alternatives because they can more credibly threaten to break off bargaining, thus affecting the other's feasible set" (Farrel 2004: 87). In exchange-theoretic terms, power spells out the "principle of least interest." Power relates to the distribution of interest and control over the resources which the actors in an exchange bargain about, and it is inversely related to the degree of dependence of one actor on the other (Esser 2000: 387f.). If A controls resources that B has an interest in, but no control over, than A is said to have *power* over B, or B is *dependent* upon A.

cannot be. In such a situation, "disparities of power are likely to give rise to mutual distrust" (Farrell 2004: 94). But if the development of trust reaches a stage where mutual identification and affect become a primary basis, then asymmetries in power are often concealed due to the shift of intentional attributions from extrinsic or instrumental to intrinsic motives (see chapter 3.1.4), and minor failures of trust will be redefined such that the available relational schemata can be maintained (Holmes 1991).

Dyadic embeddedness also allows specific expectations to be formed and stabilized, because actors can gradually learn about their dispositions, intentions, and motives ("dyadic learning"). As the relationship increases in bandwidth and richness, the actors increasingly uncover each other's characteristics and establish a basis for stable and specific expectations. On top of that, repetition fosters the development of shared relational schemata and mutual attachment, leading to "thick" affective, identification-based forms of trust. At the same time, repetition fosters the consolidation of trust into routine action (Endress 2002: 64). Empirically, a substantial body of experimental research also documents the importance of dyadic embeddedness for the development of trust (Berg et al. 1995, Buskens & Weesie 2000b, Anderhub et al. 2002, Bohnet & Huck 2004, Engle-Warnick & Slonim 2004).

Network Embeddedness describes the fact that both trustor and trustee normally interact with and maintain relationships to third parties. These can provide information or apply external sanctioning measures in response to actions taken in a trust problem. An important aspect of network embeddedness is that information about past behavior can disseminate into the networks, and also can be received from there, allowing for the emergence of *reputation mechanisms* (Kreps & Wilson 1982, Camerer & Weigelt 1988, Burt & Knez 1995, Burt 2003, Fehr et al. 2008). Reputation is an important source of trust-related knowledge and a form of social capital for the actors in question (Coleman 1988). With reputation, actors can "detach" social capital from the context of specific transactions and generalize it to other exchanges. It thus grants a certain degree of transferability. With reputation, third parties take the role of "trust intermediaries" by providing trustors with information about a potential trustee and about his or her past behavior and trustworthiness ("network learning"). At the same time, in response to the reputation-information circulating within the network, third parties can themselves reward and punish a trustee's behavior by withdrawing future trust and refusing future cooperation, by expressing social disapproval, or by inflicting otherwise costly sanctions (Burt & Knez 1995, Fehr & Fischbacher 2004). This increases the indirect costs of a failure of trust and opens up a "voice" option to the trustor, who can credibly threaten to damage reputation if the trustee does not act trustworthily. High network embeddedness also allows a trustor to more easily seek alternatives and "exit" the trust relation if the trustee is not trustworthy. Thus, network embeddedness can increase the power that a trustor has in a dyadic trust relationship. Both the threat of "exit" (by searching for alternatives) and the threat of "voice" (by

damaging reputation) change the basic structure of a trust problem ("network control"), in that additional incentives *not* to fail trust emerge (Buskens 2002).

Generally speaking, network embeddedness can induce trust even among rational and selfish actors, because different "trigger strategies" become available to ensure trustworthiness. It can completely substitute dyadic embeddedness in situations where there are many potential trustors and trustees, and an effective reputation mechanism is available (Buskens & Raub 2008). In such a situation, although there is no repeated interaction, the reputation-information that flows through a network can be sufficient to provide for favorable expectations of trustworthiness. The likelihood for trust and trustworthy response increases with network density and with the probability that information about past behavior is transmitted to other potential trustors (Coleman 1990, Buskens & Weesie 2000a). Empirically, experimental results also show that network embeddedness, in particular via reputation mechanisms, is conducive to the build-up of interpersonal trust (Camerer & Weigelt 1988, Anderhub et al. 2002, Bolton et al. 2004, Bohnet et al. 2005).

Lastly, with *institutional embeddedness* we take into account the broad cultural-normative environment, the institutions that surround a trust relation and function as a source of trust-related knowledge ("institutional learning") and as a structuring device for action ("institutional control"). An *institution* is defined here as a socially shared and sanctionable expectation with respect to the conformity to a mandatory, predescribed rule (Esser 2000c). Institutions constitute the rules of human interaction in a world of social interdependence and represent the relevance and incentive structures of a society. The incentives and sanctions provided for by institutions can be more or less formally regulated, and differ with respect to their mode of enforcement (Elster 1989, 2005). *Norms* are a class of institutions which are explicitly linked to internal or external *negative* sanctions. On the whole, institutions considerably change the opportunities and information available to actors in a trust problem (Zucker 1986, Shapiro 1987, Bachmann 1998, Ripperger 1998, Heimer 2001). We have already denoted two direct effects of institutions on trust in the form of the "structural assurance" and "situational normality" beliefs which they back up, and which contribute to the formation of expectations of trustworthiness and the willingness to be vulnerable (see chapter 2.3.1). This highlights two important aspects of institutional embeddedness: (a) institutions create a familiar background on which trust becomes possible, and (b) they provide the structural "safeguards" that enable trust between individuals in anonymous settings, even when other forms of social embeddedness are missing. Importantly, both structural assurance and situational normality are rooted in the sociological concepts of "normality" as proposed by Schütz, Garfinkel, and Luhmann, but they represent a more fine-grained distinction (McKnight & Chervany 2006).

To begin with, institutions enable trust by creating a background of familiarity. This argument is directly related to the process of interpretation, the recognition of typical things already

known and the suspension of uncertainty into a routine of unconditional trust. If stored mental schemata can successfully be applied to interpret an immediate situation (and given that this situation is not "extraordinary"), perceived situational normality will be high, yielding a sense that "everything seems in proper order" (Lewis & Weigert 1985b: 974). This enables a trustor to feel comfortable enough to rapidly form a trusting intention toward the trustee in the situation, because interactions with others are likely to occur as expected and without surprising twists (Misztal 2001). In short, "a belief in situational normality means that the people involved will act normally and can therefore be trusted" (ibid. 316). Both McKnight et al. (1998) and Misztal (2001) introduce situational normality by referring to Garfinkel's (1963) well-known crisis experiments, in which he shows that trust and the routine frames of reference quickly break down when situational normality is disturbed. All in all, this notion of situational normality closely resembles Luhmann's (1979) idea of familiarity as a precondition to trust.

In a detailed analysis, Möllering (2006a, b) carves out a theory of trust in which institutions provide the "taken-for-granted expectations that give meaning to, but cannot guarantee, their fulfillment in action" (Möllering 2006a: 363). Building on the work of Schütz (1967), as well as of Berger and Luckmann (1966) and Garfinkel (1963), he sets the "natural attitude" of the life-world as the starting point for analysis. Institutions help actors to establish the "basic rules of the game" (Garfinkel 1963: 190f.) and to maintain stable and unproblematic interaction. That is, a major function of institutions is a reduction of complexity by providing socially shared information about the likely course of action in a social context—they do so, for example, in the form of learned mental schemata about typical situations (*frame*), typical action sequences (*scripts*), typical actions by typical actors (*role*), or rules of action (*norms*, *rules*). Institutions do not simply take on the role of a third-party enforcer and guarantor—a role to which they are frequently restricted in economic accounts of trust. Instead, they must be regarded as "systems of rules and meanings that provide common expectations which *define the actors* as social beings" (Möllering 2006b: 61, emphasis added). They are not just passively consumed, but actively (re)produced in an ongoing process of symbolic interaction and "agency" (Emirbayer & Mische 1998), being both an objective fact of a socially constructed reality and an internalized part of individual identity at the same time.

When institutions instill taken-for-granted expectations, the corresponding internalized mental schemata are often enacted without question, following a "logic of appropriateness" (March & Olsen 1989). For example, actors who have internalized an institution that demands placing or honoring trust in a particular situation will do so because doing otherwise would go against their own identity and against the objective reality of society (Zucker 1986). Trust is exercised because "everybody would do so in the same position," and the actors who have internalized a relevant norm often adhere to its rule on a routine basis. According to Zucker (1986), institu-

tion-based trust derives from socially shared expectations which include, for example, symbols of membership in a group or profession, intermediary mechanisms such as contracts, guarantees, and regulations, and other sources of trust-related knowledge, such as norms and values. When the context of a trust problem indicates that certain institutions are part of the "rules of the game," this enables trust between actors because they provide the means for a social definition of the situation. Ultimately, trust and trustworthiness can themselves acquire a taken-for-granted character so that "it may be literally unthinkable to act otherwise" (Zucker 1986: 58) in a particular, familiar situation. As Kramer (1999) points out, rule-based forms of trust often trigger suspension without a conscious calculation of consequences. In essence, the sociological approach to trust suggests that institutions often routinely reduce social uncertainty and complexity for individual actors, whose main concern is how to establish shared meaning as a precondition for social action.

One can distinguish three important types of institutions integral to the notion of institution-based trust which characterize the institutional embeddedness of trust relations: (1) rules, (2) roles, and (3) routines (Möllering 2006b: 65f.). First, institution-based trust emerges and is sustained by a shared understanding regarding the system of *rules* specifying what behaviors are regarded as appropriate in a given situation. In the previous chapter, we have linked rule-based trust to moral dispositions, norms, and values when looking at the normative element in the subjective experience of trust (see chapter 2.2.3). However, the notion of rules must be apprehended much more broadly. Importantly, rules include formal law and legal contracts. *Law* represents an institution that explicitly defines sanctionable norms, and, for example in the form of contract law, very effectively reduces social uncertainty (Zucker 1986, Ripperger 1998). As Luhmann points out, "legal arrangements which lend special assurance to particular expectations, and make them sanctionable ... lessen the risk of conferring trust" (1979: 34). But instead of merely structuring action by changing the incentive structure, "contract law, trade associations and technical standards are social institutions that embody systems of rules [and meaning] for interaction," which can become "a basis for trust, if rules are understood as cultural meaning systems" (Möllering 2006b: 67). Taken together, the notion of institution-based and rule-based trust includes a broad class of "good reasons" behind expectations of trustworthiness (for example: the adherence to social norms and rule-based value systems, legal institutions such as civil law, licensing, and guarantees) and it extends the meaning and functional scope of institutions from a perspective that treats them as "external" sanctioning devices to the central role they play not only in the social definition of the situation and the reduction of social uncertainty, but also with regard to the identity of the actors participating in the social system.

Apart from rules, social roles are also regarded as an institutional basis for interpersonal trust (Barber 1983, Baier 1986, Meyerson et al. 1996). *Social roles* can be defined as sanctionable

expectations tied to a particular social position—they are a special case of a norm. Interpersonal trust enabled by social roles is "depersonalized" (Kramer 1999), because it is based on the knowledge that an actor occupies a particular social position and enacts a particular social role; it does not rest on specific knowledge of trustee characteristics. Roles evoke typical expectations concerning competence and "fiduciary responsibility" (Barber 1983), that is, the demands and obligations associated with a specific role. What is more, roles also embody typical sequences of action and typical patterns of identification and affect (Esser 2000c: 141f.). At the same time, they most directly reflect the normative and institutional structure of a society: just as different social positions are structurally related to each other (for example, via hierarchy in an organizational context), so are the social roles that individual actors fill. In this way, roles establish a fixed and expectable pattern of interpersonal relationships and interaction. They structure social positions, sanctionable expectations, and a potential course of action. Since roles are internalized during socialization, actors in fact generate interpersonal trust on the basis of their identity and self-image when shared role expectations become a basis for action (Möllering 2006a: 362). To the extent that both the intention to fulfill the role and the competence to do so are convincingly signaled by the trustee and accepted by the trustor, a trustor can choose a trusting act based on the knowledge of a normative role relation, even when dyadic or network embeddedness are absent (Buskens & Raub 2008). Taken together, social roles are conducive to interpersonal trust—a social role effectively reduces social uncertainty regarding the role occupant's intentions and abilities and thus "lessens the need for and costs of negotiating trust when interacting with others" (Kramer 1999: 678).

Lastly, an institutional basis of trust can be established from *routines*, which are "regularly and habitually performed programs of actions of procedures. They may or may not be supported by corresponding (systems of) rules and/or roles, and they represent institutions in as much as they are typified, objectified and legitimated, although their sense is mostly taken-for-granted whilst they are performed" (Möllering 2006b: 69). For example, we have introduced the notion of a *script* to denote a typical action sequence which is part of the typified and socially shared stock of knowledge. Scripts can become a basis for trust because the actors involved can take for granted that a known sequence of actions leads to expectable outcomes, while vulnerability is subjectively minimized and not greater than in past interactions (Misztal 1996). Likewise, the routine provided by modern bureaucratic institutions confers predictability in the sense that public services can be routinely and repeatedly demanded and will function "until further notice"—they therefore easily produce trust. In many cases, the choice of a trusting act and the trustworthy response become part of routine itself: when a mother "entrusts" her child to the teacher in school, she does not ask whether trust in the characteristics of the teacher is justified—her action is part of a daily routine in which doubts of this sort have been suspended. What is more, her action is embedded in an institutional environment where competence, benevolent intentions, and the personal integrity of a teacher are

based on taken-for-granted role expectations. The routinization of action is also conducive to the development of trust in ongoing relationships as it helps the actors to develop shared interpretive schemes and pass over the initial stages of a trust relation (Rempel et al. 1985, Jones & George 1998).

So far, we have examined the effect of institutions on trust through the lens of situational normality, familiarity and taken-for-grantedness, and looked at different ways in which institutions ensure unproblematic interaction to provide a basis for interpersonal trust. The second aspect of institutional embeddedness considers structural assurance—"the belief that success is likely because such contextual conditions as promises, contracts, regulations, and guarantees are in place" (McKnight et al. 1998: 478). The concept of structural assurance focuses on the sanctioning potential of institutions and their power to change the incentive structure of a trust game. It reflects a more "utilitarian" perspective on institutions, in which enforcement and deterrence become the reasoning on which a trustor can generate favorable expectations of trustworthiness. According to McKnight and Chervany (2006), structural assurance is a frequent antecedent to calculus-based forms of trust.

If, from the trustor's point of view, the effectiveness of an institution in bringing about a trustworthy response is taken for granted and its sanctioning potential is regarded as sufficient, then social uncertainty is considerably reduced: the trustor does not expect the trustee to fail trust because he knows the consequences of a failure of trust, and he can count on the effectiveness of the institution in bringing about a trustworthy response. In the words of Hardin (2001), trust is "encapsulated" in the interests of the trustee, and the trustor, by taking into account the trustee's rationale, can expect appropriate behavior. Essentially, structural assurance "may be thought of as a generalized comforting belief that reflects the effects of many types of mechanisms that support confidence in contextual actors because they provide safety-nets or prevent or redress losses due to opportunism" (McKnight & Chervany 2006: 38); it therefore grasps an important aspect of institutional embeddedness.

For example, Shapiro (1987) discusses institutional embeddedness in the form of legal contracts as a strategy for controlling the behavior of the trustee. Both parties engage in "norm making" by designing an appropriate institution in which rules, actions, and sanctions are specified. The contract changes the incentive structure of a trust problem such that the trustee does not have an incentive to fail trust, and it yields an amount of structural assurance sufficient for the choice of a trusting act, even in one-shot situations between anonymous actors. Likewise, the internalization of a social norm can be recast as the installation of an *internal* sanction mechanism which changes the structure of the trust problem such that a failure of trust has negative consequences for trustee's utility (Fehr & Schmidt 2006). If the context indicates that norm-breaking behavior will be punished, and if the trustor believes that the trustee has internalized relevant norms (including social roles), the trustor can feel "structurally

assured" and confidently expect a trustworthy response. Similarly, when reputation mechanisms are in place, they represent an institutional safeguard which delivers the structural assurance conducive to interpersonal trust. With an efficient reputation mechanism in place, a failure of trust is sanctionable and inflicts losses on the utility of the trustee. On the other hand, structural assurance can also refer to institutions that change the incentive structure with respect to the utility of the trustor. Many forms of insurance, for example, confer structural assurance insofar as they mitigate the risk of interpersonal trust for the trustor. In the case of opportunism and the failure of trust, the inflicted damage will be restored. Then, the choice of a trusting act is likely because objective vulnerability is minimized.

Note that all examples point to an important prerequisite for institution-based trust: as it is, the trustor needs to be convinced that the institution itself is effective. Problems of institution-based trust almost immediately turn our attention to the problem of system trust, which has to be solved before institutions can be an effective basis for trust development.

Overall, social embeddedness in its different variations is an integral part of a theory of trust. Since trust relations always occur in a social context, they are naturally constrained or enhanced by micro, meso, and macrolevel processes. Social embeddedness mitigates the risk of conferring trust because it provides opportunities for learning and control. Likewise, the context creates a background of familiarity in front of which the choice of a trusting act is possible. The actors have to establish a common "frame of reference" in which action can take place and be filled with meaning. In this regard, institutions take a prime role in the process of socially defining the situation; they represent taken-for-granted expectations (in the form of rules, roles, and routines) which the actors can apply to an immediate trust problem. From the perspective of situational normality, they function as cultural meaning systems that structure and control social action, while from the perspective of structural assurance, they have the power to change the incentive structure of a trust problem and enforce norm-conforming behavior—if their effectiveness is taken for granted. On top of that, institutions often provide the degree of familiarity necessary to permit the suspension of doubts on a routine basis, enabling unconditional forms of trust. In sum, social embeddedness, as a "bedrock of trust" (Shapiro 1987), enriches our understanding of interpersonal trust relations. It pins down sources of trust-related knowledge, emphasizes different mechanisms of learning and control, and highlights the important role of institutions for the generation of trust. The choice of a trusting act, from a perspective of dyadic, network, and institutional embeddedness, must be understood as a symbolic and meaningful act that relates to the context and social systems which "set the stage" for the particular trust relation.

3.2.3. Social Capital and Reciprocity

In a socially embedded trust relation, the choice of a trusting act is normally accompanied by an implicit demand of a morally correct response and a normative obligation for the trustee to prove trustworthy. When trust relations are dyadically embedded and reciprocal, the fulfillment of such an obligation constitutes a form of "asset" for the trustee. By fulfilling the obligations that come with the placement of trust, and by spending resources in the form of time, money, or cognitive effort, the trustee equally invests in a future reciprocal demand on trustworthiness. This obligation constitutes a form of *social capital* for the trustee (Ripperger 1998: 166).[12] More pointedly, in an ongoing, dyadically embedded trust relation, both trustor and trustee alternately take the role of a creditor and debtor of social capital. Likewise, network embeddedness can enable the creation of social capital in a trust relation, given that learning and control mechanisms are in place (Burt 1992, 2003). In this case, the social capital that a trustee invests in with his trustworthy response is not transaction-specific (that is, is not fixed to the particular trust relation), but his gained reputation constitutes a "generalized" form of social capital with respect to the social system (Dasgupta 1988: 175f.).

It is not surprising that trust has been a focal point of research focusing on social capital and collective action (see Lewis & Weigert 1985a). The idea that trust plays an important role in the creation of social capital and the promotion of cooperation was already endorsed by Blau, who stated that "social exchange ... entails supplying benefits that create diffuse future obligations ... Since the recipient is one who decides when and how to reciprocate for a favor, or whether to reciprocate at all, social exchange requires trusting others" (1968: 454). From a functional perspective, trust is a *mechanism for the production of social capital.* This argument is based on several observations: (1) the choice of a trusting act usually initiates a trust relation and thus constitutes an opportunity for the creation of social capital, (2) a demand on social capital warrants trustworthy action of the trustee, (3) the trustee also has to "trust" that his moral demands on social capital will be fulfilled in the future ("trust" in this sense bridges the gap between the constitution of a demand on social capital and its future realization), and (4) objectively, the true value of social capital created depends on the trustworthiness of the trustor, that is, on whether future demands of the trustee will be in fact redeemed (Ripperger 1998: 168). In short, the total amount of social capital within a social system is significantly influenced by the overall level of trust (determining the number and value of outstanding obligations) and trustworthiness (determining whether moral demands on reciprocal behavior are in fact "covered" by actual trustworthy responses).

---

[12] Ripperger defines social capital as "interpersonal obligations of a social nature, which result from a moral demand on reciprocally altruistic behavior" (1998: 166, present author's translation). In a more general notion, *social capital* is defined here as the total value of resources and services which an actor can control *via* dyadic and network embeddedness (Esser 2000b: 238, see also Bourdieu 1985, Portes 1998, Woolcock 1998).

The actors participating in social system can benefit from high levels of accumulated social capital, as it enables continual cooperation and investments that would otherwise be locked in "hold-up." However, social capital is commonly regarded as a public good (Coleman 1988, 1990: 315). It is often diminished or destroyed unintentionally because individual actors do not take into account the external effects of their actions. A trustee's individual decision about trustworthiness, if observable or available as reputation information, has an external effect on other trust relations, because the overall level of "successful" cooperation within the social system changes. In consequence, third parties are indirectly affected by the actions taken in a particular trust relation. If it is common knowledge that many participants of a social system are not trustworthy, then cooperation and the production of social capital through the mechanism of trust are severely hampered (Putnam 1993: 167). Defection undermines both trust and (future) trustworthiness, and thus the bases of social capital production.

According to Ripperger, the fragility of trust and the public good character of social capital create a "consensus to collectively control the behavior of the trustee in a trust relation and protect the stock of social capital" (1998: 184, present author's translation) with the help of social norms and other institutional measures for trust-protection. In the same line, Messick and Kramer argue that "our strong preferences for other's actions lead us to endorse and promote rules of ethics and morality, including exhortations to be trusting and trustworthy, that may be beneficial to us if we can induce others to follow these rules" (2001: 98). These authors see a possible solution to the risk of opportunism in the creation, institutionalization and internalization of social norms and rules which sustain trust and guarantee trustworthiness. Principally, these institutions aim to protect the future reciprocal demands that a trustee invests into with his trustworthy response, and thus, serve to protect the stock of social capital. In fact, one of the main propellants behind an institutional protection of trust is its beneficial effect on the production of social capital (Fukuyama 1995, Ripperger 1998f., Sztompka 1999: 105f., Messick & Kramer 2001, Burt 2003).

The most prominent example of an institution that directly relates to the function of trust as a mechanism of social capital production is the *norm of reciprocity* (Gouldner 1960), which has been identified as an almost universal norm across different cultures and different moral value systems. It directly addresses the need to "cover" outstanding reciprocal demands and ensure that social capital can be realized. The norm of reciprocity is a highly productive component of social capital production (Putnam 1993: 172), markedly decreasing transaction costs and bolstering cooperation. By far the most famous reciprocal strategy that has been examined is "tit-for-tat" (Rapoport & Chammah 1965), but the specific reciprocal norms that individuals learn vary significantly from culture to culture and across different types of situations (Ostrom 2003). As a consequence of early socialization, actors tend to reciprocate each other's behavior in an almost reflexive way because they have internalized the rule, and social sanctions are

almost universally applied to violators (Allison & Messick 1990). Importantly, if actors have fully internalized the norm of reciprocity (or any other norm of trust-protection, for that matter), then motivations for action do no longer lie in the instrumental consequences on utility, but in an intrinsic value that emerges from norm-compliance, and in the form of a "bad conscience" which inflicts negative psychological costs in the case of a failure of trust (Elster 2005: 202f.). In a social system in which compliance to the norm of reciprocity is "the rule," social exchanges can be more easily established because structural assurance is high, and rule-based forms of trust and trustworthiness can be favorably expected. That is, the norm of reciprocity functions not only to stabilize social relationships, but also as a "starting mechanism" to initiate social interactions and trust (Gouldner 1960). This boosts cooperation and the creation of social capital, and, coincidentally, allows trust relations to mature to advanced developmental stages in which a reliance on norm-compliance may even be no longer necessary and be replaced by shared routines and mutual identification. An effective norm of reciprocity thus is a prime example of a successful institutionalization of a rule to protect and maintain trust and trustworthiness (Ferrin 2008).

3.2.4. Trust and Culture

The institutionalization of trust is discussed by sociologists mainly in a historic perspective, and it is usually linked to the fundamental question of social order, which was first raised in the course of industrialization and modernization (Misztal 1996). For example, Durkheim (1984, [1893]) criticized "atomistic" social contract theories by showing that the noncontractual part of the contract, that is, the unspoken "et cetera assumptions," qualifications, and provisions for future action are backed up by society as a "silent partner," through which contracts as a social institution become viable. This noncontractual aspect of contracts and agreements is largely based on trust (Collins 1982: 12). To act with "good faith" in agreements, promises, and contracts means that "et cetera assumptions" are respected and taken for granted. Otherwise, and from a position of distrust, any form of commitment would always appear incomplete and cooperation would be prevented by the insurmountable risks of opportunism. Durkheim developed this argument in light of his concept of "organic solidarity," according to which a moral consensus, based on the recognition of an increased interdependence (resulting from the division of labor) is a central source of integration and solidarity, even in modern societies.

Parsons added to the idea that trust is an indispensable ingredient for the maintenance of social order by introducing the concept of *generalized media of symbolic interaction* (commitment, influence, money, and power) as a basis for interaction, cooperation, and social integration (Parsons 1967, 1971). With the concept of a social media of exchange, he suggests the principal "channels" which structure, control, and sanction individual action and facilitate the continuous reproduction of social systems. Trust is regarded as a central foundation of these

media. According to Parsons, trust in their reliability, effectiveness and legitimacy is a primary condition for their functioning (Parsons 1963: 46ff.). In addition to that, trust is required to bridge unavoidable "competence gaps" (Parsons 1978: 46) between experts and lay-persons during professional interactions, which he regards as a key aspect of modernity and a product of increasing structural differentiation and specialization. In line with Durkheim, Parsons holds that the bases of trust lie in shared normative orientations: "People defined as sharing one's values or concrete goals and in whose competence and integrity one has confidence come to be thought of as trustworthy individuals or types" (Parsons 1978: 47).

Luhmann (1979) proposes that the transition from small and undifferentiated societies into modern technologically and organizationally complex social structures is paralleled by changes in the types and functions of trust which are necessary to integrate them. As pointed out before, he distinguishes between interpersonal and system trust. Importantly, Luhmann emphasizes that the functioning of modern societies is less and less dependent on interpersonal trust, while system trust is becoming increasingly important—especially with respect to legitimacy of bureaucratic sanctions, safeguards, and the legal system (ibid. 48). A standard argument to underpin this view is that group size and other group-related attributes can drastically influence the effectiveness of social norms and the success of attempts to institutionalize them (Olson 1965, Hardin 1982). In short, learning and control mechanisms to detect and punish defectors can be more easily established in small groups, where there is also less scope for free riders to profit, and where cooperative efforts are more directly targeted towards specific individuals and outcomes. Likewise, Zucker (1986: 11f.) asserts, with regard to rule-based forms of trust such as social norms, that they may back up "local" forms of enforcement, while "global" environments and larger social systems need other foundations. Thus, in modern societies and large-scale market-based economies, it becomes increasingly difficult to establish trust based on institutions that function best in small-scale environments. Consequentially, "local" mechanisms have to be replaced or complemented by other forms of institutional protection, which largely depend on system trust.

In this line, Giddens states that modernity is marked by "disembedding, i.e. the 'lifting out' of social relations from local contexts of interaction and their restructuring across indefinite spans of time-space" (1990: 21). This is achieved by the use of symbolically generalized media of exchange (e.g. Parsons, Luhmann) and expert-knowledge systems, which also serve as "access points" to reembed complex social systems in concrete interactions and particular trust relations (for example, in the form of a patient-physician relationship). From this perspective, trust is integral to modern society because it is the mechanism that bridges the gaps in time and space and enables the reembedding of social systems via access points—effectively, "all disembedding mechanisms ... depend on trust" (ibid. 26). In line with Luhmann, Giddens argues that there is an increased need for trust in modern societies. But he em-

phasizes that the increased demand for trust pertains to both interpersonal and system trust simultaneously (e.g. system trust in the medical system and rule-based interpersonal trust in the physician), opposing Luhmann's assertion that interpersonal trust becomes less important.

In consequence, modern societies are typically marked by a build-up of *institutional frameworks* ("trust settings," "rounding frameworks of trust," ibid. 35) to protect and maintain trust and the functioning of the system of society—for example, in the form of bureaucratic regulations, standardization, professional ethics, legal sanctions, and insurance. As Shapiro puts it, "in complex societies in which agency relationships are indispensable, opportunities for agent abuse sometimes irresistible, and the ability to specify and enforce substantive norms governing the outcomes of agency action nearly impossible, a spiraling evolution of procedural norms, structural constraints, and insurance-like arrangements seems inevitable" (1987: 649).

The evolving mix of local and global mechanisms for the protection and maintenance of trust is frequently analyzed from a macrolevel perspective. Taken together, different institutional measures for trust protection, the different norms, prevalent cultural practices, and legal safeguards merge into a unique *trust culture*: "Trust culture [...] is a system of rules—norms and values—regulating granting trust and meeting, returning and reciprocating trust; in short, rules about trust and trustworthiness" (Sztompka 1999: 99, see also Fukuyama 1995). A society's trust culture circumscribes the totality of cultural and normative-institutional rules which concern trust and trustworthiness, while presenting themselves as social facts *sui generis* and as properties of the social system (or, as Lewis & Weigert 1985b put it, as a "social reality"). As we have seen, these rules can stem from moral values and rule-based value systems (honesty, benevolence, integrity etc.), from diverse role expectations and shared social norms (reciprocity, truth-telling, keeping secrets, being fair, etc.), from cultural practices (for example, the general rule of *noblesse oblige*, demanding exemplary conduct from those who have attained elevated positions in the social hierarchy), and a plurality of "normalizing" institutions which enable rule-based form of trust; they may also describe more diffuse expectations pertaining to trust and distrust, such as stereotypes and prejudices.[13] Once a culture of trust emerges and becomes ingrained in the normative system of society, it can become a vital factor influencing both the choice of a trusting act and its trustworthy response. In an established "positive" culture of trust, "people not only routinely tend to, but are culturally encouraged to express a trustful orientation toward their society, its regime and institutions, fellow citizens, as well as their own life-chances and biographical perspectives" (Sztompka 1998: 21).

[13] "In cultures of trust, some rules may be very general, demanding diffuse trustfulness toward a variety of objects, and expressing a kind of certitude about the good intentions of others, implied by overall existential security. There may also be more specific rules, indicating concrete objects as targets of normatively demanded trust or distrust. Object-specific cultural trust or distrust is often embedded in stereotypes and prejudices [...] There are also culturally diffuse rules demanding and enforcing general trustworthiness" (Sztompka 1999: 68f.).

The concept of trust culture can be regarded as a continuation of a branch of political research focusing on the links and causal interrelations between culture and democracy. It is closely related to the ideas of "civic culture" (Almond & Verba 1972) and the "civil society" (Seligman 1997), and, paralleling the works of several other political trust researchers (Fukuyama 1995, Putnam 1995), primarily represents a theoretical attempt to outline the cultural preconditions for the functioning of modern democratic institutions. The emergence of a trust culture is characterized as a continuous process in which choices about trust and trustworthiness, influenced by surrounding social structures and the preexisting climate of trust, generate trust-confirming or trust-disconfirming events. These experiences, normally widespread and socially shared, cumulate and turn into routine, and eventually into normative rules (Sztompka 1999: 119f.). In effect, positive experiences of trust furnish the development of a culture of trust; negative experiences will eventually generate a culture of distrust, or a "culture of cynicism" (Putnam 1995). Notably, the trajectories of cultural development are "self-amplifying," and can result in "virtuous loops" or "vicious loops," depending on whether trust-confirming or trust-disconfirming events prevail (Sztompka 1999: 120). That is, if trust is usually honored, the process moves toward building a culture of trust, whereas failed trust pushes development toward suspicion, which can damage even an established trust culture.

Sztompka (1999: 122ff.) identifies five structural factors that determine the direction of cultural evolution: (1) First, "normative coherence," that is, a solid normative ordering of social life which raises a "feeling of existential security and certainty" (ibid. 122), is seen as encouraging trust and the development of trust culture. In our terminology, this refers to an aspect of high situational normality. Importantly, normative coherence means that trust-related norms (e.g. demands for honesty, loyalty and reciprocity) are effective and regarded as sanctionable, indicating what people will and should do, and making behavior predictable in ordered, unproblematic "fixed scenarios" (ibid.). Sztompka contrasts this to a state of anomy, in which social rules and norm enforcement are "in disarray." As a result of low situational normality, the perceived uncertainty and insecurity widely increases, pushing the cultural development towards a climate of distrust.

(2) Second, the "stability" of the social structures at large—that is, whether social institutions, networks, associations, organizations, political regimes, and so forth are "long-lasting, persistent and continuous" (ibid.) or change rapidly—will influence the development of trust culture. Continuity provides a basis for routinization and lends security and comfort in trust relations. The choice of a trusting act and a trustworthy response then easily become a matter of habit, whereas rapid social change (for example, in the case of revolutions) undermines situational normality—in phases of quick social change, "nothing is certain anymore," and prevalent norms, social roles, everyday routines, and habitualized patterns of action may no longer be adequate, raising feelings of "estrangement, insecurity, and uneasiness" (ibid. 123). This

increases the probability that trusting expectations will not be met and that trustees will not respond as expected, triggering suspiciousness and the tendency to withhold trust.

(3) Furthermore, the "transparency" of social organization—that is, whether information about the functioning, efficiency, and levels of achievement (as well as failures and pathologies of social institutions) is available or not—is regarded as an important factor (ibid.). Transparency effects pertain to an aspect of system trust—if principles of institutional operation and the *modus operandi* of social systems are visible, then even failures or dysfunctions of the social system do not necessarily come as a surprise to actors. In this sense, transparency allows actors to "relate" to the social systems, assuring them about what can be expected. On the other hand, if principles of operation are vague and hidden, then a general climate of suspicion and distrust may emerge and hamper the choice of trusting acts, undermining a culture of trust.

(4) Another factor is "familiarity" with the environment of the trust relation, which "breeds trust," and produces a "trust-generating atmosphere" (ibid. 124) in which expectations of trustworthiness become favorable and confident. Sztompka links this directly to the "natural attitude" of the life world, providing security, certainty, and predictability, whereas in situations of "strangeness," actors react with anxiety, suspicion, and distrust. Sztompka develops this point for the case of migrants and migrant communities, citing a classical study of Thomas and Znaniecki (1927) about Polish emigrants in the United States who suffered a great deal from unfamiliarity with their new environment, which raised a culture of distrust.

(5) Lastly, "accountability," that is, the presence of formal or informal agencies monitoring and sanctioning the conduct of a trustee, is regarded as conducive to the build-up of a positive trust culture. The concept of accountability can be directly restated in terms of structural assurance introduced earlier (see chapter 2.3.1): "Accountability enhances trustworthiness because it changes the trustee's calculation of interest, it adds an extra incentive to be trustworthy, namely to avoid censure and punishment" (Sztompka 1999: 88). Thus, when functioning institutions provide efficient control, the risk of opportunism and defection is decreased, and confident expectations of trustworthiness can be formed: "Everybody is confident that standards will be observed, departures prevented, and that even if abuse occurs it will be corrected by recourse to litigation, arbitration, restitution, or similar. This stimulates a more trustful orientation toward others" (ibid. 125).

On the whole, we can use the concept of "trust culture" to grasp the overall social conditions prevailing in a given society which are conducive or disruptive to the build-up of interpersonal trust. Most accounts stress the positive side-effects of an existing culture of trust and the successful institutionalization of trust-related rules, roles, and routines: it increases "spontaneous sociability" (Fukuyama 1995: 27f.), "civic engagement" (Almond & Verba 1972: 228), is

regarded as highly productive component of social capital (Putnam 1993, 1995), and, overall as an "integrative mechanism that creates and sustains solidarity in social relationships and systems" (Barber 1983: 21). A well-established trust culture is frequently regarded as indispensable and as a desirable "good" in itself: "A nation's well-being, as well as its ability to compete, is conditioned by a single, pervasive cultural characteristic: the level of trust inherent in a society" (Fukuyama 1995: 7). However, although positive consequences of an institutionalization of trust are preferably accounted for, a strong culture of trust can also lead to undesirable consequences—for example, social closure and corruption—as Gambetta (1993) shows in working out the relevance of a culture of trust for the success of the Sicilian mafia. Whether a culture of trust should be regarded as a "good" in itself is a normative question which will not be subject of further analysis here.

The most important result of the preceding analysis is that there exist a number of socially prescribed interpretive schemes or "trust settings" (Giddens 1990) to encourage trust on a cultural basis. A "culture of trust" may emerge as a consequence of the co-evolution of local and global mechanisms for trust protection, which at the same time are a primary means for the social integration of modern societies. If trust becomes institutionalized, it is commonly produced on the basis of "how things are done" (Zucker 1986: 12); that is, based on habitualized cultural practices and routinely executed social roles and norms. The successful (re)production and maintenance of trust culture can be traced back to individual socialization and to the learning and internalization of relevant mental schemata. After all, it is the shared mental models and interpretive schemes that carry the "culture of trust." Their application provides a ground for the development of rule-based interpersonal trust, following a "logic of appropriateness," in which the cultural-normative context of situations guides the choice of a trusting act from the pillars of situational normality and structural assurance.

Empirically, the study of cultural differences in trust is one of the most flourishing areas of current trust research (see Saunders et al. 2010 for an extensive review). Scholars scrutinize how trust relations develop within and across cultural boundaries, how the preconditions to trust differ between cultural domains, and how trust can be maintained in cross-cultural contexts. Overall, these studies, far too numerous to be reviewed in detail, lend considerable support to a perspective that acknowledges the influence of socialized and learned cultural models and the prevalent "trust culture" on the practical development of trust. For example, a number of studies have demonstrated differences in the observed levels of trust and trustworthiness between different cultural domains (for example, "individualist" *versus* "collectivist" societies). It has also been found that culture influences the production of trust-related cues in relationships, and, by providing the "interpretive lens" used during interaction, serves as a filter for the signals emitted from others. Trust may be backed up by the recognition of (shared) cultural identity, but cultural boundaries may also become a barrier to trust when the context

is becoming increasingly unfamiliar to the actors. Some researchers have argued that the nature and quality of trust vary greatly over different cultural domains, and that, accordingly, the meaning of trust also differs. One important methodological conclusion that can be drawn is that trust research must take into account the cultural idiosyncrasies and peculiarities stemming from the prevalent trust culture when conducting empirical research. The influence of "culture" on trust is considerable, and the continuing empirical support provided by intercultural trust research suggests that any broad conceptualization of trust must take into account its cultural roots, and refer to the learned stock of trust-related knowledge that defines the scope and extent of trust.

3.3. The Economics of Trust

3.3.1. The Rational Choice Paradigm

The economic perspective on interpersonal trust marks a "current mainstream" (Möllering 2006b: 13) and "major approach" (Bigley & Pearce 1998: 411) to the phenomenon of trust, linking it to the paradigm and theory of rational choice (e.g. Gambetta 1988b, Coleman 1990, Cook 2001, Hardin 2002, Ostrom & Walker 2003). The abstract simplicity of economic models allows a formal representation of complex ideas in a clear and parsimonious way, and highlights the "logic" of decisions in situations with a well-defined objective structure, making rational choice approaches the "most influential images of trust" (Kramer 1999: 572) in contemporary research. Simply put, interpersonal (dis)trust is warranted or withheld by rational utility-maximizing actors who, in face of constraints and directed by their preferences, goals, and incentives, have to make a decision about the choice of a trusting act.

But what is a "rational" and "utility-maximizing" decision? How are "preferences" and "incentives" represented, and how do actors come to make a choice? Before we can proceed to highlight the ways in which interpersonal trust is modeled, it is important to circumscribe the theory of rational choice and to pin down fundamental assumptions of the rationalist paradigm. Notably, "rational choice theory" is not a unified theoretical framework, but the term subsumes under its umbrella a number of different variants (such as expected utility theory, game theory, evolutionary economics, and marginal analysis). Yet, all variants have some methodological and theoretical considerations in common which are characteristic of the rationalist paradigm. The following postulates can be characterized as the "hard core" of the rational choice research program (see Elster 1986a, Hedström & Swedberg 1996, Esser 1999b: 295f., Opp 1999, Boudon 2003, Gintis 2007, Kirchgässner 2008: 12ff.).

First, rational choice theory rests on the principle of *methodological individualism.* This principle holds that social phenomena must be explained in terms of individual actions (Coleman 1990: 11f., Esser 1999a: 91f.). The main unit of analysis is the single actor; collective phe-

nomena such as cooperation, the production of a public good, the conclusion of a contract, or the functioning of "perfect markets" are analyzed and explained from an individual perspective. Methodological individualism can be distinguished into a "strong" and a "weak" version, which differ with respect to the way that social aggregates and collective phenomena are treated in the explanans (Hedström & Swedberg 1996). While the strong version does not accept any references to aggregates, the more prominent position, which also informs the present work, is a position of weak methodological individualism, which accepts that not all elements in the explanans of a scientific explanation need to be dissolved into the individual-level components—for reasons of tractability and "for the sake of realism" (ibid. 131). For example, effective social norms (a product of individual action) need not be explained again in terms of third party compliance when it can be convincingly argued that they objectively shape the constraints an actor faces when defining a situation.

Second, explanations in the rationalist paradigm are explicitly *analytical* and *intentional*. Any rational choice explanation proceeds by first constructing a model of the situation to be analyzed. In doing so, only the essential elements are abstracted from the problem at hand. Thus, the final object of analysis is an analytical abstraction of reality, representing the vehicle of explanation. Of course, the model is incomplete—but to the extent that it captures the "essential" ingredients, it will shed light on the real world situation that it is intended to explain. On top of that, human action is assumed to be intentional and principally understandable; explanations recur to the intentions of actors in explaining the choice of action. "An intentional explanation [...] seeks to provide an answer to the question of why actors act the way they do; and to explain an action intentionally means that we explain the action with reference to the future state it was intended to bring about" (Hedström & Swedberg 1996: 132). This allows the researcher to "understand" action in the sense postulated by Weber. As Coleman notes, "Rational actions of individuals have a unique attractiveness as the basis for social theory. If an institution or a social process can be accounted for in terms of the rational actions of individuals, then and only then can we say that it has been 'explained'. The very concept of rational action is a conception of action that is 'understandable', action that we need ask no more questions about" (Coleman 1986: 1).

Third, every actor is assumed to have clear *preferences* which motivate concrete behavioral goals, define the actor's interests and allow him to direct his behavior towards alternatives of choice. Preferences are the fundamental source of motivation and must satisfy certain conditions in order to be suitable for modeling rational action (von Neumann & Morgenstern 1944). Importantly, they must be *complete*, meaning that all alternatives can be compared pairwise and brought into a preference relation, and *transitive*, meaning that no logical errors occur in the full preference relation that includes all alternatives. Furthermore, preference must satisfy the principle of *independence*, requiring that a preference relation between two alternatives

must not be distorted by introducing a third one. Lastly, preferences are assumed to be *continuous*, so that preference orderings cannot be lexicographic. These assumptions of "choice consistency" are fundamental for the rationalist paradigm because they ensure that preferences can be represented by a numerical function to evaluate the alternatives. An actor is supposed to act according to his own preferences only, and not according to the preferences of others. Of course, his preferences may take into account the interests of others, and thus an actor may come to act benevolently, altruistically, or malevolently; his preferences may also include prosocial orientations that shift his goals away from "self-interest seeking with guile" (Williamson 1975: 9). But the "axiom of self-interest" is normally supposed: the actor acts in accordance with his own preferences. Preferences reflect the actor's idea of value as they have been developed during socialization (Esser 1999b: 359f., Kirchgässner 2008: 12).

Fourth, every decision problem contains *opportunities* and *restrictions*. Opportunities come in the form of a set of alternatives among which the actor can choose. It is not necessary that "all possible" alternatives are known to the actor, but the alternative-set must be fixed in a given decision problem. Any alternative is connected to some course of action and a number of resulting consequences. In addition, certain restrictions limit the freedom of choice and the scope of action (that is, they narrow down the alternative-set to a "feasible set"). For example, the income of an actor, the market prices of goods or the legal framework are objectively "given" and cannot be changed, ruling out certain alternatives. A decision-maker has to respect these "material" constraints, but the scope of action may also be limited by a number of "social" constraints, for example by social norms and institutions which prevent or proscribe certain courses of action. Generally, the environment is characterized as being subject to scarcity: resources such as time, money, energy, and so forth, are not available in unlimited amounts, which means that certain actions, although desirable, cannot be executed because material, cognitive, or physical resources are lacking.

Fifth, the actor possesses *information* about the choice situation. This information may be perfect or imperfect (see chapter 2.2.2 already). With perfect or "full" information, the actor knows his preferences and he can also determine precisely the consequences of each alternative. If other actors are involved and the situation is one of strategic choice, then his information includes the knowledge of their preferences as well. On the other hand, if information is not perfect, an actor will have to make a decision based on his *beliefs* about possible action opportunities and their effects. If these beliefs pertain to future events and states, they are normally labeled "expectations"; these may be unambiguous or ambiguous. Thus, imperfect information introduces an aspect of risk or ambiguity, which pertains not only to future states (i.e. the consequences of a course of action) but also to the preferences of other actors involved, so that an actor does not know for sure which preferences and intentions other actors have. This mirrors the aspect of irreducible social uncertainty resulting from the imperfect

knowledge of other's preferences and the corresponding intentions and motivations, which we have already discussed. Preferences, constraints, and information are the basic ingredients of a model of rational action.

Sixth, given preferences, restrictions, and information, an actor evaluates the different alternatives at his disposal, taking into account the costs and benefits of each alternative, weighing the pros and cons of the consequences, and finally choosing an action. Preferences define the actor's interests, and actions serve the purpose of fulfilling these interests (the principle of instrumentalism). If preferences are consistent, they can be expressed by a numerical function (a "utility function") which the actor uses to evaluate the alternatives and to determine their utility. The core nomological assumption of the rational choice framework is the *principle of utility maximization*: an actor chooses that alternative which best satisfies his interest (Elster 1986a). That is, actors maximize their utility subject to the beliefs and the constraints they face. We can say that an action is *rational* if it satisfies the principle of utility maximization under constraints.[14] Rationality in this sense means that, given preferences and restrictions, the actor is able to determine the course of action which he prefers to all others or, at least, to determine those courses of action which he prefers and those about which he is indifferent. With these assumptions, rational action appears "reasonable" or "appropriate" to the extent that, given the constraints and available information, the actor chooses a course of action which best serves his own interest. The principle of utility maximization represents the basic rule of choice on which the "logic of selection" rests in the rational choice paradigm. Note that an actor's decision is ultimately directed by the expected consequences of action (the principle of consequentialism).

This behavioral model is the nomological core of the rational choice paradigm, commonly identified by the assumption that preferences and constraints affect behavior and that individuals in some way maximize. A formal specification of the postulates is given by expected utility (EU) theory and its close relative, subjective expected utility (SEU) theory (see Schoemaker 1982, Mas-Colell et al. 1995: 167f.). EU theory is rooted in the core assumptions presented above, but it specifies more precisely how preferences and restrictions are causally linked to action, and thereby formulates a concrete choice rule. It reveals the relation between the independent variables "expectations," "evaluations," and the dependent variable "choice of action." Given a set of alternatives $A = (A_1, ... A_x)$ and consequences $C = (C_1, ... C_y)$, an actor evaluates the consequences according to his utility function, so that $U(C) = (U_1, ... U_y)$. The information available is expressed in the form of probabilities $P = (p_{11}, ... p_{xy})$ which denote the likelihood that a certain consequence will eventuate, given that a specific alternative

[14] The axioms of transitivity and completeness ensure that a decision will maximize the utility of an actor (if he follows his preferences), and are therefore sufficient to induce rationality in the above sense (Mas-Collel et al. 1995: 6).

was chosen. The expected utility of an alternative A_i then can be found by weighing the utility-evaluated consequences with their probability of occurrence so that

$$EU(A_i) = p_{i1}*U_1 + p_{i2}*U_2 + ... + p_{iy} * U_y = \sum p_{(i)}U_{(i)}$$

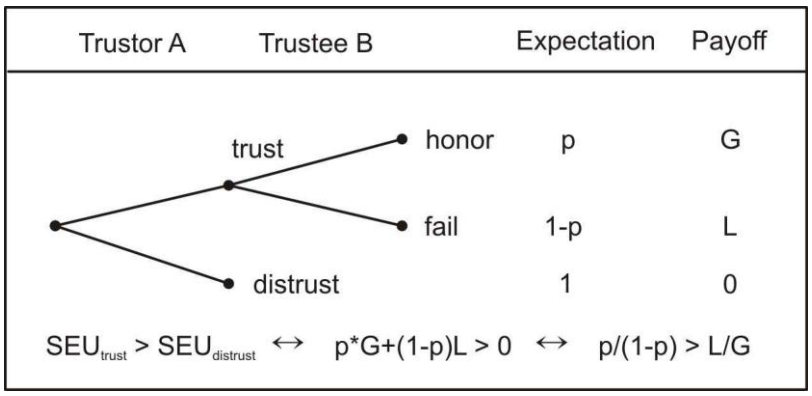
The principle of utility maximization demands that the alternative with the highest expected utility is chosen. While EU theory assumes that the objective probabilities of the occurrence of events are known to the actors, SEU theory emphasizes that the available information about events is often limited, or deviates from the objectively true value. To accommodate for this fact, objective probabilities are replaced with subjective counterparts by means of a transforming function $P = w(p)$, so that $SEU(A_i) = \sum w(p_{(i)})U_{(i)}$, and expected utility is represented by subjective expected utility (Edwards 1954, Savage 1954).

3.3.2. Modeling Trust

If the choice of a trusting act is to be represented formally, then all variables which influence the trustor's decision must be identified and included in a model of the decision process. An intuitive formal model of the choice of a trusting act was given by Coleman (1990: 91f.) in his conception of trust as a binary choice under risk. Note that trust, in rational choice models, is defined in terms of observable, instrumental choice behavior. The model will serve as a starting point for our further exploration of the "economics of trust."

Coleman states that trust problems are special cases of the more general class of decisions under risk: "The elements confronting the potential trustor are nothing more or less than the considerations a rational actor applies in deciding whether to place a bet ... If the *chance of winning*, relative to the *chance of losing*, is greater than the *amount that would be lost* (if he loses), relative to the *amount that would be won* (if he wins), then by placing the bet he has an expected gain; and if he is rational, he should place it" (ibid. 99). Coleman proposes formally grasping all relevant aspects of a trust problem in three variables: First, the potential gains G, relative to the status quo, which may be obtained in the case of a trustworthy response. Second, the potential loss L, relative to the status quo, which would be incurred if the trustee were not trustworthy; and third, the subjective probability p, which represents the trustor's subjective estimate of the probability that a trustworthy response occurs. In effect, all trust-related knowledge is represented by the single expectation of trustworthiness, p. Both G and L represent the trustor's evaluation of the consequences (see figure 8).

Figure 8: Coleman´s trust model



Following logic of SEU, and assuming that exactly two alternatives (trust or distrust) exist, Coleman points out that a rational trustor chooses a trusting act if SEU(trust) > SEU(distrust), that is, if p/(1-p) > L/G. In essence, the model hypothesizes a threshold value for p, defined in relation to G and L, which is sufficient to induce the choice of a trusting act. This corresponds to Gambetta's (1988a) idea of trust as a threshold value to which the actual expectations are compared. The idea is also indicative of the core of the rationalist paradigm: a trustor will rationally trust a trustee if he perceives a net expected gain. As Gambetta notes, optimal threshold values will vary subjectively as a result of individual dispositions, and will change with situational circumstances (represented here by G and L). Once a favorable expectation exceeds the threshold, the actor will engage in risk taking behavior and choose a trusting act. Coleman suspects that every individual has a standard estimate of p, accruing to situations in which one deals with strangers (deriving from dispositional tendencies and generalized expectations), although p can be replaced by specific expectations p^+ in close relationships, which are normally higher than their generalized counterpart (Coleman 1990: 104). This redraws the distinction of generalized and specific expectations made by Rotter.

Some authors argue that, in order to speak of trust "proper," it is necessary that the potential losses involved exceed the potential gains, so that L>G (Deutsch 1958, Luhmann 1979: 24). This implies that p > 0.5 and establishes a special requirement in order to interpret an expectation as favorable.[15] Only if the subjective probability of a trustworthy response is greater than the subjective probability of a breach of trust, will the actor engage in choosing a trusting act. Coleman explicitly rejects such a position, maintaining that trust is similar to a bet in which the alternative with a higher subjective expected utility is chosen. In his view, social embeddedness and institutional mechanisms (such as repetition, reputation, and social norms) are the most important sources of trust-related knowledge, because they enable effective incentive

[15] L>G implies L = G+x, x>0; so that pG > (1-p)L ↔ pG > (1-p)(G+x) → p > 0.5 (G+x)/ (G+0.5x). It follows that p >0.5, since (G+x)/(G+0.5x) > 1.

mechanisms to protect and safeguard trustworthiness, especially in close communities with intact communication structures and a high flow of information (Coleman 1990: 100, 108f.).

It is worth noticing the high level of abstraction of Coleman's model. It formulates conditions which allow an outside observer to interpret certain choices of action as trustful. His conception of trust aims at an explanation of choice behavior. While the choice of a trusting act is the explanandum, the underlying rational choice principles (SEU theory) represent the means to understand and causally explain it. Scrutinizing the model for the role of information and knowledge, it is apparent that trust is crucially dependent on the trustor's subjective expectation p. Coleman notes that, in many situations, p, L and G are known with varying degrees of certainty, and further states that p is often least known, which is why actors should engage in a search for information. This "will continue so long as the cost of an additional increment of information is less than the benefit it is expected to bring" (ibid. 104). Yet, ultimately, trust hinges on the "fixed" model variables, and, centrally, on the (unambiguous!) expectation of trustworthiness. As Harvey (2002b: 291) notes: "In the language of economics, trust can be viewed as an expectation, and it pertains to circumstances in which agents take risky actions in environments characterized by uncertainty or informational incompleteness." The perspective on trust taken by rational choice advocates therefore is an articulately cognitive one, and Hardin aptly notes that "my assessment of your trustworthiness in a particular context *is simply my trust* of you" (Hardin 2002: 10, emphasis added, see also Gambetta 1988: 217). Note that this implies a special causal relation between expectations and trust: expectations and evaluations *explain* trusting behavior, and thus are causal antecedents.

The abstractness of the decision problem modeled by Coleman results from the fact that the model variables are not further specified; they are assumed to condense all experience and the knowledge of the trustor, as well as his evaluations in a particular trust problem, reflecting the individual's history of learning and socialization as well as perceptions of current situational constraints or opportunities. The model displays the interdependencies between the basic variables which influence choice, but it leaves open the question of their emergence and formation in a specific situation—preferences, the alternative-set and information must be "fixed" in order to satisfy the axioms of rational choice. Although Coleman accepts recourse to specific and generalized expectations, and admits to processes of information search, his perspective does not allow a conception of trust in which interpretation and prereflective processes play any further role. Evaluations and expectations are given, and they define trust completely. The notion of trust as a mechanism for the reduction of social complexity thus receives a very unique reading: trust is not a "reason" for a particular course of action or a process that fosters to the formation of favorable expectations; it is merely a characteristic of action, which results, along the way, if the constellation of the model's components is accordingly favorable.

A reduction of complexity, or a notable "suspension" of uncertainty, must have already taken place.

3.3.3. Encapsulated Interest

Hardin (1993) remarks that rational choice approaches to interpersonal trust are characterized by two central elements: the first element is the trust-related knowledge of the trustor, expressed in his expectations. Expectations assume a prominent role in almost all economic models of interpersonal trust (Hardin 2003: 81). Although the trustee's preferences and intentions are private information and can never be known with certainty, the trustor can condense his knowledge into an estimate of how likely a trustee will respond in a trustworthy manner, using the different categories of trust-related knowledge. However, as Hardin (1993: 153) laments, many models of trust—including the one of Coleman—only implicitly refer to the second fundamental element: the actual incentives of the trustee to be trustworthy and to fulfill the trust. They are, as Hardin claims, equally important to any rational choice account of trust: "you trust someone if you have adequate reason to believe it will be in that person's interest to be trustworthy in the relevant way at the relevant time [...] one's trust turns not on one's own interests, but on the interests of the trusted. It is encapsulated in one's own judgment" (Hardin 1993: 152f.). Therefore, a rational trustor must "take a look at the world from his [the trustee's] perspective as it is likely to be when it comes to his having to fulfill his part of the agreement" (Dasgupta 1988: 51). Taking the perspective of the trustee is reasonable, as the trustor must assume that the trustee is an equally rational, utility-maximizing actor. In contrast to a narrow contemplation of self-interest and "bald expectations" (Hardin 2003: 83), the encapsulated-interest account of trust advises a more sophisticated understanding of the trustee's interests, opportunities, and constraints.

With respect to the relative neglect of the trustee's rationale, Hardin points out that, "surprisingly, much of the literature on trust hardly mentions trustworthiness, even though implicitly much of it is primarily about trustworthiness, not about trust" (2002: 29). In order to model the choice of a trusting act, we first have to understand the decision of the trustee, assuming an equal amount of rationality, and then to "encapsulate" it in the trustor's decision. Therefore, Hardin goes on to argue, the trustee's rationale is indeed of primary concern for trust research. The encapsulated-interest account suggests that the trustee's choice must be analyzed with equal scrutiny, and it shifts our focus towards a simultaneous consideration of both actors (and decisions) involved in the trust problem.

One step in this direction is to include the trustee as a second actor into the economic models. The "parametric" decision problem of the trustor, as formulated by Coleman, is then recast in

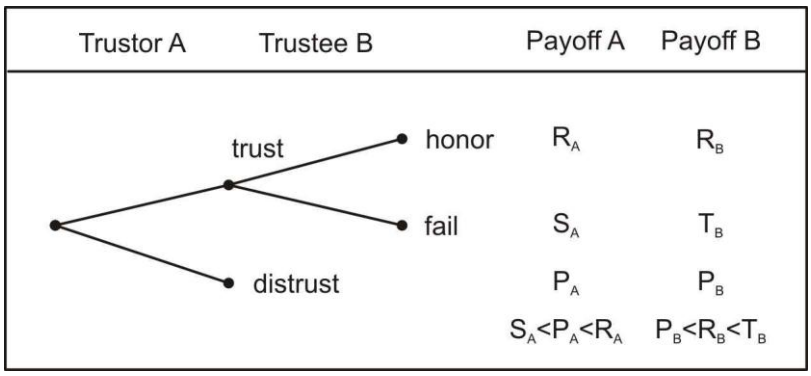
terms of strategic interaction between trustor and trustee. These types of problems can be analyzed using the apparatus of game theory (see Fudenberg & Tirole 1991, Gintis 2000b).[16] Unsurprisingly, game theory represents a major stream of research in the rational choice approach to trust (James 2002b, Camerer 2003, Buskens & Raub 2008).

Game theoretic models are used to analyze situations of strategic interdependence, problems of cooperation or coordination, and social dilemmas. An "extensive form game" consists of a tree-like structure with nodes. Each node indicates which player can make a move at that node. A move consists of an action that a player can take at a node, choosing from a set of actions belonging to the node. At the end-nodes of the game tree, the player's payoffs are indicated. A general assumption is that of *common knowledge*: the game is known to each actor, each actor knows that it is known to each actor, and so forth. Game theory is unexceptionally concerned with the question of *equilibrium*, that is, whether and which outcome(s) of an interaction, given the strategic situation, can be rationally expected. The assumption of player rationality means that each player engages with the goal of maximizing expected utility. The solution to a game theoretic problem is described in terms of players' *strategies*, which specify what each player would do at each decision node of the game. Player's strategies must satisfy certain properties that put constraints on what a rational actor should do. In the famous "Nash Equilibrium" (Nash 1950), each player chooses a strategy that is a best response and maximizes his expected payoff, given the strategies of all other actors. In Nash equilibrium, no player has an incentive to deviate from his strategy, given that the other players play their equilibrium strategy. In this sense, Nash equilibrium behavior is the basic game-theoretic specification of individual rationality. The concept of "Subgame-Perfect Equilibrium" (Selten 1965, 1975) is an equilibrium refinement which rules out irrational behavior off the equilibrium path. Subgame-perfect equilibrium consists of strategies which form a Nash equilibrium for the game and also for each subgame (that is, for each part of the game tree which can be considered a proper game tree and thus forms a subgame).

The most basic game that can be used to model interpersonal trust is known as the "trust game" (Camerer & Weigelt 1988, Dasgupta 1988, Kreps 1990). It was informally introduced in chapter 2.1.2 when describing the basic trust problem. In trust research, the trust game is considered a benchmark scenario; it resembles a one-shot interaction between two actors A (the trustor) and B (the trustee) and it contains all the structural ingredients of the basic trust problem but a more general formulation of payoffs (figure 9):

[16] The following short introduction is necessarily brief, keeping an emphasis on understanding and intuition rather than on mathematical precision with respect to terminology, proof of theorems, and modeling approaches to trust.

Figure 9: The trust game



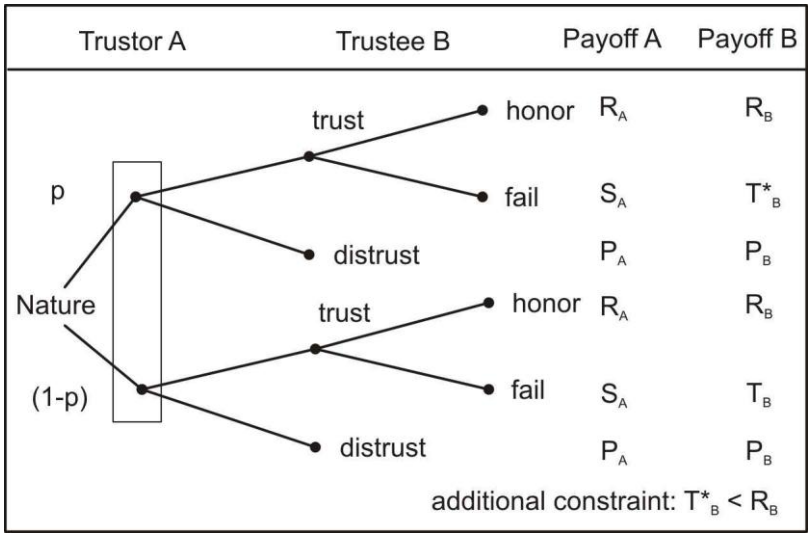
The *status quo* payoffs are represented as the pareto-inefficient "punishment" outcomes P_A, and P_B. Honored trust yields "reward" payoffs R_A and R_B, and failed trust means that the trustor receives the "sucker" payoff S_A, while the trustee can gain the "temptation" payoff T_B. The game is fully described by adding the following payoff relations: $S_A<P_A<R_A$, that is, the trustee receives a net gain from honored trust, but incurs a loss from a failure of trust, and $P_B<R_B<T_B$, that is, the trustee has an incentive to fulfill trust (there are mutual gains from trust/trustworthiness), but there is also a temptation to defect.

The game has a unique solution in terms of each player's strategies. The only subgame-perfect Nash-equilibrium that exists is for the trustor to always distrust and the trustee to always fail trust. The surprising prediction from game theory, and the "paradox" solution presented by the benchmark scenario, is that a rational trustor would never choose a trusting act, because a rational trustee would always fail trust.[17] The trust game is therefore a classical example of a "social dilemma," in which pareto-efficient collective outcomes are prevented by individual rationality (Ostrom 1998, 2003). Of course, this prediction crucially hinges on the assumed payoff structure of the game. If the trustee's payoffs for being trustworthy somehow were larger than his payoff for failing trust, the equilibrium prediction would be a combination of strategies in which trust and trustworthiness prevailed. As suggested by the encapsulated-interest account, incentives for the trustee to fulfill trust may lead to a more efficient equilibrium. But the benchmark scenario neglects potential effects stemming from social embeddedness. There are no histories, no reputation, and neither institutions nor social norms influence the purely self-interest actors. It is a "raw" game devoid of the social, institutional and cultural context.

[17] This kind of forward reasoning is introduced by "backward induction," which is necessary to establish subgame-perfection: starting from the terminal nodes and working backwards toward the start, each best move at each node is determined. Since the trustee would choose to fail trust in order to gain the payoff T_B, we can reduce the decision problem of the trustor to a choice between the "safe" alternatives P_A and S_A. Since we have $P_A > S_A$, the trustor will rationally choose to distrust.

In the above model, an explicit assumption is that the players are perfectly and completely informed about all aspects of the game—an assumption which is at odds with real-life situations, where information about preferences, motivations, and utility is private. To make the trust game more realistic and to model the aspect of asymmetric information and social uncertainty, we can introduce imperfect information in the sense that the trustor does not know which type of trustee he will meet (Harsanyi 1967, 1968). The trustee can either be a trustworthy or an untrustworthy type. The trustor is only informed about the probability p of a random move of nature which determines the trustee´s type at the beginning of the game. In the picture shown below, the trustee is trustworthy with a probability of p, since then $R_B > T_B^*$; he is not trustworthy with a probability of (1-p, in which case $T_B > R_B$ and the trustee always fails trust (see figure 10).

Figure 10: Trust game with incomplete information



The subgame-perfect equilibrium of this refined game is straightforward to identify. Choosing the alternative of distrust yields the expected *status quo* payoff EU(distrust) = P_A. By choosing a trusting act, the trustor´s expected utility is EU(trust) = p R_A + (1-p) S_A, so that a rational trustor would choose a trusting act if $P_A < p * R_A + (1-p) S_A$. Rearranging terms yields the following equilibrium condition for trust: $p > (P_A-S_A) / (R_A-S_A)$. Note that this equilibrium solution of the two-player game coincides with Coleman´s formulation of the choice of a trusting act, once we reinterpret the random move of nature as the trustor´s subjective expectation of trustworthiness p. Generally, most economic models can be easily extended to incorporate imperfect information, but the derivation of equilibria and formal analysis become increasingly complex. In the following, we will stick to games with complete and perfect information for ease of demonstration.

A voluminous body of empirical evidence suggests that the benchmark prediction of the standard trust game is quite pessimistic in comparison to what actors actually do in real-life situations, or in behavioral experiments, where trust and reciprocity are much more frequently observed (see James 2002b, Ostrom & Walker 2003, Fehr & Schmidt 2006, Johnson & Mislin 2011). In order to explain the discrepancy between theoretical predictions and empirical results, the basic trust game has been modified in many alternative ways to incorporate incentive effects stemming from dyadic, network, and institutional embeddedness. The common element of these modifications is that they change the incentives in such a way that it will not be rational for a trustee to exploit trust (James 2002b).

A very prominent example is to bring an aspect of history and dyadic embeddedness into the game by repeating the stage game (Axelrod 1984, Kreps 1990, Gibbons 2001, Anderhub et al. 2002, Bicchieri et al. 2004). As noted before, dyadic embeddedness enables mutual learning and control, and thus changes the way in which trustor and trustee will reason about the game. The possibility of repeating successful interactions creates a mutual interest in continuation, and serves as a means of encapsulating the interests of the other party. At the same time, unwanted sequences of play can be punished by exiting the trust relation and threatening to refuse future cooperation. This allows for more complex strategies that include contingent decisions in each round based on the outcomes of previous rounds. The trustee has to counterbalance his short-term interests of failing trust with his expected future pay-offs, and the incentives may change in such a way that a trustworthy response can be rationally expected. Formally, if the stage-game is repeated a number of times t, with probability $v \varepsilon [0,1]$ after each round, then the present expected value of some repeated outcome X equals to $EU(X,t) = X + v * X + ... + v^{t-1} * X + v^t * X = X / (1 – v)$ in the limiting case of an indefinitely repeated game, where v represents the "shadow of the future" (Axelrod 1984). The larger v, the more important future outcomes become. It can be easily verified that trust and trustworthiness are optimal strategies in the indefinitely repeated trust game if $v \geq (T_B - R_B) / (T_B – P_B)$.[18] Note that the equilibrium condition is independent of the trustor's payoffs—it refers solely to the trustee's rationale, formally restating an argument of encapsulated interest.

The equilibrium is subgame-perfect and relies on the trustor playing a "grim trigger strategy." In this strategy, the trustor will distrust for the rest of the repeated game once the trustee has failed trust. The result is based on several implicit assumptions: first, it is assumed that the threat of exit is credible. Second, there is common knowledge of the game and its payoff-structure. And lastly, it is implicitly assumed that all actions can be perfectly monitored with-

[18] The trustee can expect $EU(honor) = R_B / (1-v)$ if he is always trustworthy. A one-time defection with subsequent distrust yields the net discounted payoff $EU(fail) = T_B + v * P_B / (1-v)$. The trustee will honor trust if EU(honor)>EU(fail), which, after rearranging, yields the condition $v \geq (T_B - R_B) / (T_B – P_B)$. Since we have $P_B<R_B<T_B$, the right-hand side represents the "temptation" of the trustee to defect, relative to the reward from honoring trust and the *status quo*.

out additional monitoring costs. Even if these assumptions appear rather stringent, the equilibrium condition above proves that trust and trustworthiness can be supported in a repeated game between rational, selfish actors. In fact, following the "folk theorem" for repeated games (Fudenberg & Maskin 1986, Fudenberg & Tirole 1991: 150f.), there exist a large number of such equilibria, including the reciprocal tit-for-tat strategy (Axelrod 1984) and other trigger strategies in which the trustor withdraws trust only for a limited number of times. This form of conditional cooperation is also termed "weak reciprocity" in the sense that it is supported and can be accounted for in terms of self-interest (Gintis 2000c). Empirically, repetition fosters trust both when the trust game is played with limited and unlimited time-horizon (Engle-Warnick & Slonim 2004, 2006).

Dyadic embeddedness and the emerging trigger strategies which sustain trust can be interpreted as a form of dyadic control: in close resemblance to reputation within social networks, information about past behavior is evaluated and influences future contingent action. If both trustor and trustee are embedded into social networks via third-party relationships, then network learning and control, by which information about past behavior is transmitted to other potential trustors, can take an influential role in establishing trust and trustworthiness (Burt & Knez 1995, Bohnet et al. 2005). While dyadic embeddedness necessitates that the two parties meet again in the future, this need not be the case with network embeddedness and reputation—here, interactions are typically one-shot or marked by uncertainty as to whether the interaction partner stays the same (e.g. a temporary buyer-seller relationship on eBay). However, "historic" information about the trustee's past behavior is available to the trustor when making a choice in the stage game. The folk theorem and the trigger strategy argument presented above extend straightforwardly to these situations: a reputation mechanism may provide the incentives for a rational trustee to induce trustworthiness, because other potential trustors can refuse future cooperation and withhold trust if the trustee's reputation indicates that he is not trustworthy and has previously failed trust. The related economic models become rather complex and need not be spelled out in detail here (see Raub & Weesie 1990, Buskens & Weesie 2000a, Buskens 2003), but they provide a solid economic underpinning of reputation and network embeddedness to the development of interpersonal trust.

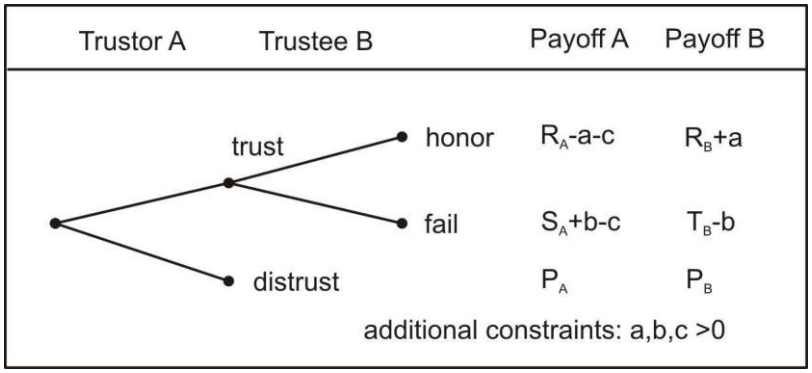
3.3.4. Contracts and Agency

While both repetition and reputation rely on some sort of "history" of play, several other refinements and alterations have been proposed and incorporated into the trust game to account for the fact that, even in one-shot situations with neither a dyadic history nor available reputation information, trust and trustworthiness are regularly higher than predicted by the benchmark scenario. These variants add a number of institutional solutions to ensure that a rational trustee can be induced to act trustworthily, and a rational trustor can be motivated to trust—showcasing once more the importance of the encapsulated-interest notion of trust to economic

modeling. The most commonly analyzed institutions are binding contracts, hostage posting or other forms of "credible commitment," and punishment and sanctioning mechanisms, or combinations thereof. All in all, these solutions address aspects of institutional embeddedness.

As with repetition and reputation, binding contractual agreements between the trustor and the trustee modify the incentives of the trust game in such a way as to make the trustworthy option preferable to the trustee (Malhotra & Murnighan 2002, Colombo & Merzoni 2006, Ben-Ner & Putterman 2009). A very simple modification for the one-shot benchmark scenario would be a penalty b>0 that the trustee incurs if he fails trust. A second possibility would be some form of additional reward a>0 for being trustworthy. Furthermore, it is assumed that the contract has some transaction cost c>0 to the trustor, which include negotiation and monitoring, as well the costs of enforcement (James 2002). The following figure shows the payoff modifications to the trust game for a contract which includes both punishment and reward, and is costly to implement (figure 11):

Figure 11: Trust game with contracts

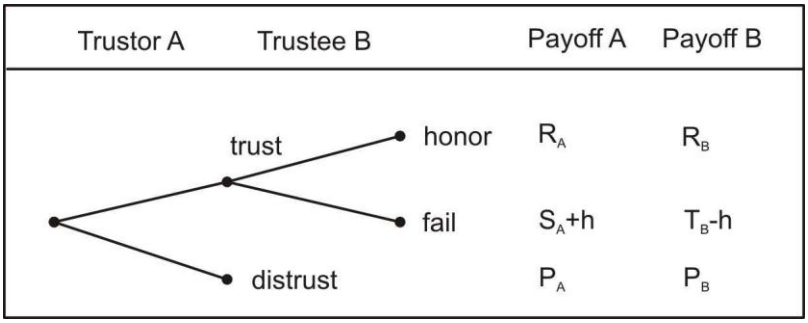


In this situation, the trustee will sign the contract and honor trust if $a+b > T_B-R_B$, that is, whenever the contractual incentives compensate the opportunity costs from not failing trust. A rational trustor will be willing to negotiate the costly contract as long as $R_A-P_A > a+c$, that is, whenever the monitoring costs and the rewards that need to be paid to B do not exceed the increase in wealth relative to the *status quo*. If the parameters of the contract are fixed accordingly, then the contract is sufficient to induce a pareto-efficient equilibrium in which trust and trustworthiness prevail.

Another institutional solution frequently proposed considers instances where the trustee commits himself to a trustworthy response (Weesie & Raub 1996, Raub 2004, Bracht & Feltovich 2008, Servatka et al. 2011) by means of a preplay decision in which he invests in a credible signal in order to communicate trustworthiness. Depending on the specific way such commitment is modeled, the trustor puts his future utility into "escrow," either by directly transferring some amount of his income to the trustor, so that the trustor can keep it if trust is failed, or by

investing into a "hostage" that he loses if he fails trust. The hostage may or may not be redressed to the trustor to compensate his losses. In any case, the hostage is a "sunk cost" that cannot be recovered (one example of such a hostage would be a product guarantee that a manufacturer give to its products). Consider the simple case where the hostage is a "sunk cost" and not redressed (see Bracht & Feltovich 2008). If the trustee chooses to post a hostage of value h, then the reduced subgame that results after the precommitment stage would include the following payoffs payoffs (figure 12):

Figure 12: Trust game with pre-commitment and hostage posting



Thus, the trustee's response depends on the size of the hostage. It can be easily seen that the hostage serves as a credible signal to commit to trustworthiness if it is large enough. In the example, the hostage binds the trustee if $h > T_B - R_B$, so that it exceeds the potential gain from failing trust. Raub (2004) develops a more complicated model which includes imperfect information with respect to the type of the trustee, as well as uncertainty with respect to exogenous events (contingencies that may bring about an unfavorable outcome irrespective of the trustee's actual choice of action). He explicitly models the preplay stage in which the trustee can choose whether to post a hostage or not, and derives necessary and sufficient conditions for a "pooling equilibrium," in which trustworthy and untrustworthy types of trustees use the hostage (it is thus not a reliable signal of trustworthiness), and "separating equilibrium," in which only trustworthy types use the signal (in which case it is reliable). Importantly, he derives upper and lower bounds on the value of the hostage for it to be used by trustees and accepted as a credible signal by trustors in the different equilibrium conditions. The hostage-value is dependent upon the amount of exogenous and endogenous social uncertainty. A closely related form of "commitment" that is beneficial to the buildup of trust is gift-giving (Camerer 1988).

Lastly, the trust game can be modified to include options for punishment and sanctioning of the trustee by first parties (Bohnet & Baytelman 2007), by third parties (Fehr & Fischbacher 2004, Charness et al. 2008), or by some external device which ensures that a failure of trust has a substantial cost to the trustee. The models closely resemble the formal solutions presented above—essentially, incentives and the utility function of the trustee are modified such that

a trustworthy response can be "rationally" expected. Empirically, punishment options are regularly found to be effective as a means to ensure trustworthiness (Fehr & Gächter 2000a, 2002a, Houser et al. 2008, Mulder 2008, but see the discussion below).

A theoretical framework that can be used to further explore the intricacies of economic solutions to trust problems is principal-agent theory (PAT, see Mas-Colell et al. 1995: 477f.). Generally speaking, PAT is concerned with situations in which one actor (the principal) hires another actor (the agent) to perform a task which will bring him some gain in return ("agency"). The principal's gain is, however, directly related to the agent's performance. The agent has to make an effort and incurs costs to perform the task, but he also receives a reward that is linked to his performance, or paid as a fixed wage. Being a utility-maximizing rational actor, the agent would ideally like to "shirk" instead of "work" in order to minimize his effort and yet still earn the income as specified in the contract. The principal, on the other hand, suffers from limited information concerning the agent's skill and preferences, and he cannot directly monitor the agent's performance once the contract has been signed. That is, he faces substantial social uncertainty and vulnerability with respect to the fulfillment of the contractual obligations. All in all, the "raw" structure of the principal-agent-problem closely resembles a basic trust problem (Shapiro 1987, Ensminger 2001): the principal is in the positions of a trustor who must decide whether to trust the agent with respect to his characteristics and motivation to perform. The agent is in the position of a trustee, who may or may not fulfill the content of the trust relation (the task) and has an informational advantage concerning his skill and preferences. Therefore, "trust in an agency relationship means that the principal perceives the agent to be motivated to put in the full effort required to produce the principal's benefit and to justify the agent's reward, even though opportunism cannot strictly be ruled out" (Möllering 2006b: 32).

A core tenet of PAT is that whenever an individual engages another individual to whom some decision-making authority is given via a transfer of control, a potential agency problem exists, and agency costs (that is, transaction costs in the form of signaling and monitoring costs) are incurred, diminishing overall welfare. The agency problems discussed within PAT pertain to asymmetric information before concluding a contract (adverse selection), the motivational problem of the agent (moral hazard), and stagnation due to potential investments that the principal might have to undertake before concluding the agreement, which, if their return is uncertain, might undermine any contractual engagements in advance (hold-up). To overcome the problem of adverse selection, moral hazard and hold-up, PAT proposes different solutions ("mechanism design"), whereby an efficient solution to the problem of agency must minimize the agency costs (Jensen & Meckling 1976). The problem-set of PAT can straightforwardly be transferred to problems of interpersonal trust (Ripperger 1998: 63f., Heimer 2001, James 2002a).

The PAT framework can be used to inform trust research in two ways: First, it suggests that contracts and other incentive mechanisms are suitable for backing up institution-based forms of trust, as outlined above. Contracts and legal arrangements covered by enforceable punishment opportunities can be regarded as a suitable "foundation" for further trust development, and initiate legitimate forms of institution-based trust (Shapiro 1987, Lorenz 1999). Second, trust relationships can be interpreted as an informal agency relation, which means that the apparatus of PAT can be used to analyze trust in terms of *implicit psychological contracts* (Rousseau 1989, Robinson 1996, Ripperger 1998). In this perspective, trust represents a cost-free solution to the problem of agency, rather than its basic problem; this perspective vividly explicates the celebrated notion of trust as a social "lubricant" (Dasgupta 1988: 49) governing transactions and reducing transaction costs.

Explicit contracts are not unequivocally accepted as a solution to the problem of trust. In the literature, it is debated whether trust and control are mutually exclusive (Das & Teng 1998, 2001, Möllering 2005b).[19] As some researchers claim, the organization of monitoring and sanctioning procedures is complex, and the very existence of formal control mechanisms has the potential to undermine trust by creating an atmosphere of distrust (Fehr et al. 1997, Ostrom 2000, Fehr & Gächter 2002b, Hardin 2002: 127). For example, contractual relations may require overt calculation so that the involved risks become salient. The unwanted side-effects of contracts can stem from monitoring activities, threat, or litigation. These actions may evoke conflict, opportunism, and more defensive responses. Put sharply, "trust is not a control mechanism, but substitutes for control ... People do not need to develop trust when their exchange is highly structured and easily monitored ... Some controls actually appear to signal the absence of trust and, therefore, can hamper its emergence" (Rousseau et al. 1998: 399).

Empirical studies have provided evidence that contractually safeguarded exchange relationships tend to "crowd out" intrinsically motivated trust (Frey & Jegen 2001, Malhotra & Murnighan 2002, Fehr & List 2004, Mulder et al. 2006). As Malhotra and Murnighan (2002) point out, explicit contracts lead the parties involved in the trust relation to attribute their behavior to the contract ("situational attribution") rather than to favorable characteristics of the other ("dispositional attribution"). Since contracts present a very salient situational feature, and since behavior will be regarded as strongly regulated, "contractually mandated cooperation may provide an insufficient basis for continued cooperation if contracts are no longer available ... Someone who has only been known to cooperate under the constraints of a binding con-

[19] For example, Möllering claims that "trust and control each assume the existence of the other, refer to each other and create each other, but remain irreducible to each other" (2005: 283). In a similar fashion, Noteboom asserts that "trust and control are substitutes in that with more trust there can be less control, but they are also complements, in that usually trust and contract are combined, since neither can be complete" (2006: 247).

tract, might not, in the absence of the contract, be expected to cooperate because he or she is not seen as trustworthy" (ibid. 538). In contrast, informal mechanisms such as promises and assurances ("nonbinding contracts"), by their very nonrestrictiveness, should lead to positive dispositional attributions if the exchanges are successful, and therefore allow for the development of favorable specific expectations of trustworthiness. On top of that, a contractual solution to a trust problem must always be a second best solution in terms of efficiency, because it has an agency cost to it, which decreases the trustor's net benefit and overall welfare.

Do contractual agreements and more generally, all other forms of institutionally safeguarded trust, in fact *not* belong to the phenomenon of interpersonal trust, then? Clearly, the answer to this question depends on our definitional choices. As suggested at the very beginning of this work, a basic trust problem is marked by social uncertainty and vulnerability, and the choice of a trusting act is frequently explained by recurrence to favorable expectations of trustworthiness. Superficially, one could argue that in situations of high institutional regulation, the basic trust problem is nonexistent because incentives change in a way that makes trustworthy responses certain, thereby removing a core aspect of the trust problem. What is more, trustors obviously do not develop specific expectations about trustee characteristics once they can base their choices on structural assurance, as the empirical evidence suggests. When the institutional safe-guards are removed, observed levels of trust are lower as compared to dyads where trust has been developing on an informal basis. This adds to the "crowding-out" argument, and seems to invalidate institutional safeguards as a basis for trust.

Yet, for the aim of developing a broad conceptualization of interpersonal trust, it is irrelevant which source of trust-related knowledge constitutes the starting ground for favorable expectations in the particular instance. Note that a contract, or any other institution for that matter, does neither remove the trustee's principal option to fail trust, nor the social uncertainty inherent in the trust problem, even when economic modeling suggests that it does. This claim hinges on the implicit assumption that system trust and structural assurance are close to perfect, and questions of enforcement are not an issue. The fact that institutions are mostly "taken for granted," aptly recognized as familiar, confidently expected to regulate behavior, and thus breeding institutional trust, does not, on a theoretical level, make them less valid factors for a causal explanation. On the contrary, the much more interesting question arises how actors achieve the conditions that foster institution-based trust, given that they can never be "certain" about the effectiveness of an institution. In fact, when institutional (for example, legal) enforcement is uncertain, as is the case in countries with high political instability and weak enforcement, low levels of system trust can prevent even the seemingly unproblematic contract-based solution to the problem of interpersonal trust (Sztompka 1996, Bohnet et al. 2001). Once institutions are taken into account, we have to address the aspect of system trust and structural assurance (Luhmann 1979, McKnight et al. 1998), which shifts the problem of in-

terpersonal trust to a "second-order" problem of system trust as a basis for institutional trust. In economic models, it is implicitly assumed that institutional back-ups of trust are actually enforceable.

While the problem of agency is classically solved by designing appropriate institutional mechanisms that rely on formal, extrinsic, monetary incentives to control the behavior of the trustee, the psychological interpretation of PAT suggests their replacement with trust as an informal mechanism, which, in combination with reliance on internal incentives, sanctions and rewards, constitutes an "implicit" agency relationship (Robinson 1996, Ripperger 1998). Basically, accepting the trustor's investment constitutes a form of *implicit psychological contract* between the trustor and trustee, in which the rights and obligations accruing to the content of the trust relation are determined (Rousseau 1989, 1995). This includes sanctionable expectations of trustworthiness stemming from the placement of trust as well as the acquisition of moral demands on future trustworthiness by the trustee. The main distinction between explicit and implicit contracts is that the latter are not directly enforceable by third parties—only the parties involved can determine whether an agreement has been violated and pursue its enforcement. But similarly to the explicit approach, this requires that action be directed by appropriate incentives. The psychological variant of PAT therefore turns to internal sanctioning and reward to explain intrinsically motivated decisions. To account for such motivational sources in corresponding economic models, it is necessary to extend the standard apparatus of economic theory. Preferences must include "soft factors," such as the internalization of norms, altruism, fairness considerations, and feelings of guilt—that is, they must include the social preferences of actors.

3.3.5. Social Preferences

Repetition, reputation, contracts, and external punishment are examples of how the interests of the trustor and the trustee can become encapsulated by embedding the trust game in a social context. Apart from external incentives and sources of motivation, internal incentives also influence the trustee and trustor's decisions. For example, a trustee might have internalized a social norm which he feels compelled to honor, or might be guided in his behavior by moral principles that he values highly: both of these possibilities are likely to impact his judgment and choice of action. Likewise, certain standards of fairness or justice may motivate the trustee while reasoning about trustworthiness, and would have to be taken into account when modeling the trustworthy response. A trustee might try to judge whether a reciprocal response is "justified," and he might feel guilty when failing trust, fear the repercussions of breaking a social norm, or derive some intrinsic utility from being a "good-doer."

Empirically, a large number of experiments have called into question the exclusive focus of economic models on material self-interest (Fehr & Fischbacher 2002, Fehr & Gintis 2007). It

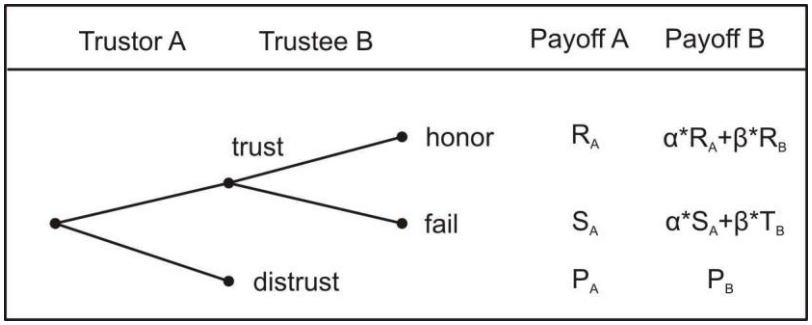
seems as though actors respect not only their own well-being, but also that of other actors involved in the exchange. Presumably, this is one prime reason why hypotheses of standard rational choice models, such as the benchmark scenario trust game, are regularly rejected by empirical data. The existence of *social preferences* implies that the utility of an actor depends, in some way or the other, on the utility of the interaction partners, on their actions, their intentions, or on the process of interaction (see Kolm & Ythier 2006 for an extensive review). In short, social, other-regarding preferences cause an internalization of external effects of action, which is sometimes interpreted as "bad conscience," and equated with internal mental costs and rewards (Rilling et al. 2002, Fehr & Schmidt 2006, Fehr & Camerer 2007). Recent neuroscience studies suggest that social preferences play an important role in trust problems (Zak 2004, Fehr et al. 2005, Zak 2005, Baumgartner et al. 2008, Fehr 2008), a finding that is also supported by data from behavioral experiments (Cox 2002, Bohnet & Zeckhauser 2004, Cox 2004). Social preferences are activated genuinely in social contexts, and differ from preferences to take unsocial risks even on a neural basis (Fehr 2008). Thus the simple analogy of trust "as a subcategory of risk" does not hold once we regard the neural processes involved.

To model social preferences in an economic model, it is necessary to extend the apparatus of standard game theory to include psychological factors (Geneakoplos et al. 1989, Battigalli & Dufwenberg 2009). Importantly, in a *psychological game*, the utility of an actor not only depends on the outcomes of the game but also on his beliefs before, during or after play. This is captured by assuming belief-dependent preferences, combined with an assumption of rational Bayesian belief-updating at every node in the game tree. In addition, psychological games allow for *higher-order beliefs* ("A believes that B believes that ...") that can capture belief-dependent motivations, intentionality, and, more generally, conjectures about other actor's states of mind. The equilibrium concept used to analyze psychological games is "psychological sequential equilibrium," a sophisticated refinement of subgame perfection. Informally speaking, in a psychological sequential equilibrium, actors play strategies which are optimal given their beliefs, and they hold beliefs which are optimal and turn out to be true, given the strategies played (the requirement of consistency of player's assessments).

In the simple case of altruistic preferences, an actor will evaluate the outcome of an interaction depending on the payoffs that other actors receive, in addition to his own payoffs (Andreoni 1990, Levine 1998). *Altruism* implies a positive correlation between the utility of an actor and the utility-level of those actors who are influenced by his actions. Therefore, altruistic preferences motivate action independently of external constraints. The satisfaction of altruistic preferences represents an intrinsic incentive for trustworthy behavior: the trustee then is motivated to act in a trustworthy manner because he gains additional utility from doing good. In the economic framework, this incentive exists as long as the additional utility from altruistic action compensates the opportunity costs from not failing trust. More concretely, assume that

an altruistic trustee B has the following utility function: $U_B = \alpha * A(X) + \beta * B(X)$, where A(X) and B(X) denote the material payoffs of the trustor and the trustee at an end-node of the decision tree. The weights α and β determine the relative importance of each actor's payoff to the utility of the trustee. The ratio α/β can be interpreted as the *social orientation* of the actor (Lahno 2002: 63). The actor is egoistic and strictly maximizes his own utility if α is zero, and he is purely altruistic or completely socially oriented if β is zero (figure 13):

Figure 13: Trust game with altruistic preferences



Assuming these preferences and applying the model to the trust game, a trustee is trustworthy whenever $\alpha/\beta > (T_B–R_B) / (R_A–S_A)$. In this case, the trustee's expected utility from altruistic trustworthiness is larger than the expected utility from failing trust, because the welfare-level of the trustor is taken into account. The trustor's choice of action, when accounting for the trustee's rationale, thus depends on his belief about the social orientation of the trustee. This can be captured in a standard trust game with imperfect information.

A closely related class of models considers preferences of *fairness* and *inequity aversion* (Fehr & Schmidt 1999, Bolton & Ockenfels 2000). The main difference in modeling is that the outcomes are now evaluated against some normative standard of equity. For example, in the Fehr and Schmidt (1999) model, players experience disutility whenever they perceive an outcome as inequitable; that is, whenever they are worse off relative to some reference point, or whenever other players are worse off relative to some reference point. The relative standing of the players is included in the preferences as a potential disutility that accrues whenever the outcomes deviate from the normative default. Thus, fairness considerations and inequity aversion can motivate a trustworthy response if the trustee perceives that a failure of trust would put the trustor into a disadvantage.

Such models of altruistic and equity-oriented preferences can explain trustworthy behavior. Therefore, they appear to be a suitable vehicle for an explanation of interpersonal trust in terms of "encapsulated interest." However, they are incompatible with a bulk of empirical evidence showing that individuals are regularly prone to punish others for their behavior, even if punishment is costly and does not yield any additional material payoffs (Fehr & Gächter

2000a, 2002a, Gintis et al. 2003, Fehr & Fischbacher 2004, Charness et al. 2008). In other words, individuals frequently choose actions that do not suggest a completely unconditional interest in the utility of others, or a mere concern for distributional fairness. What is more, when assuming preferences of the above kind, then "only outcomes matter"—that is, actors have preferences over the ex-post distribution of wealth, but they do not evaluate the process by which the final outcomes have been arrived at (Falk et al. 2008).[20]

In contrast, empirical data suggest that humans elicit a kind of *strong reciprocity*, in the sense that both reward and retaliation can be intrinsically motivated and triggered by the actions of others, independent of the immediate impact on material payoffs. That is, actors are not unanimously motivated by concern for the utility of others, but their concern is dependent on what other actors choose to do given the circumstances. Such a possibility opens up when we assume that actors condition their choice of action and the evaluation of final outcomes by the intentions ascribed to others. This was already suggested by Gouldner, who asserted that the norm of reciprocity is not unconditional: "To suggest that a norm of reciprocity is universal is *not*, of course, to assert that it is unconditional ... obligations of repayment are contingent upon the imputed *value* of the benefit received ... the *resources* of the donor ... *the motives imputed to the donor* ... and the nature of *constraints* which are perceived to exist or be absent" (Gouldner 1960: 171, emphasis added).

A model of such intention-based strong reciprocity was proposed, for example, by Falk and Fischbacher (2006).[21] In this model, reciprocity is dependent on how kind or unkind an action is perceived to be, relative to some evaluative standard; and perceived kindness triggers a reciprocal response, given the available strategies. Actors judge the kindness of an action at every decision node by reasoning about the intentions which have potentially motivated the observable action, and by evaluating their inference relative to some fairness standard, such as equity. Furthermore, they assess the influence of their own actions on the utility of others, as expressed in a measure of reciprocation. The product of the terms "kindness" and "reciprocation" enters the utility function of the actor as an additional utility that models strong reciprocity; it is weighted with person-specific parameter τ, so that $U_B = B(X) + \tau$ * kindness * reciprocation (notation and representation are simplified here to capture the essentials of the model). The parameter τ can be interpreted as a result of individual socialization; it captures how strongly the norm of reciprocity has been internalized. The size of τ determines the relative weight of reciprocal motivations in comparison to purely material self-interest B(X). If an ac-

[20] With regard to a wide class of equity-models of fairness (all of which are based on preferences over outcome distributions, just as the altruism model presented above), Falk et al. note that "recently developed inequity aversion models ... are incomplete because the neglect fairness intentions" (2008: 289) and, documenting further empirical evidence, argue for the importance of intentions.

[21] Related models were proposed by Rabin (1993), Charness & Rabin (2002), Dufwenberg & Kirchsteiger (2004). See Fehr & Schmidt (2006) for a discussion.

tor attributes kind intentions to the other player and sees opportunities to increase the other's utility (the product term is positive), then he can maximize his own utility with positive reciprocal behavior. Likewise, if an action if perceived to be unkind, and if punishment opportunities which diminish the other actor's utility exist (the product term is again positive), then an actor can maximize his utility with negative reciprocal responses.

In a reciprocity equilibrium, the actors choose actions that (1) are best responses in the sense of a subgame perfect Nash-equilibrium, and (2) are consistent with the initial beliefs that prove to be correct during play (Falk & Fischbacher 2006: 302)—the reciprocity equilibrium is a psychological equilibrium. Importantly, note that reciprocity is in fact motivated by social preferences, and not by expected future payoffs, as in the case of retaliatory tit-for-tat and other conditionally cooperative strategies. Thus, even in one-shot situations without dyadic embeddedness and a history of play, actors will react to kind and unkind actions, and can increase their utility by choosing an "appropriate" response. With respect to the problem of interpersonal trust, the model predicts that a trustor's choice of a trusting act depends on his beliefs concerning the reciprocity parameter τ of the trustee. Trustworthiness, when motivated by strong reciprocity, can increase the trustee's utility, and its probability depends on the size of the investment made by the trustor (Falk & Fischbacher 2000). The result is intuitive: the probability of a trustworthy response increases with the "importance" of the trust relation. A trustee will feel more compelled to honor trust placed in him if there is a large risk involved and "much is at stake" for the trustor. However, the model does not suggest a differentiation between actors—the inclination to act reciprocally is the same with friends or strangers. There is also no further reference to the normative-cultural context in which the particular trust problem is embedded.

The approach showcases the intrinsic value of action to the actors when complying with a norm of reciprocity. In the language of economics, it is the opportunity cost of a foregone utility gain that the actors incur if, given attributed intentions, they do not choose a reciprocal response. Importantly, compliance to the norm of reciprocity can be optimal and utility maximizing; the intrinsic motivation does not enter the model in the form of an anticipated punishment cost. As pointed out in chapter 3.2.2., honoring trust also constitutes a moral demand on future reciprocation since trustworthy responses are usually connected to some form of cost and effort. Apart from the direct effect of norm-compliance on utility, this adds an indirect incentive to respond trustworthily.

Considering the empirical evidence, intention-based reciprocity models have received considerable support from experiments—however, their applicability does not universally extend to all domains of social life: apparently, these models work well in the domain of "revenge" and negative strong reciprocity, whereas the support for "reward" triggered by kindness and positive strong reciprocity is rather mixed (Fehr & Schmidt 2006). This is particularly daunting

with respect to the problem of explaining interpersonal trust in one-shot situations, where we cannot fall back to arguments involving weak reciprocity and dyadic embeddedness as a means to encapsulate the interests of the trustor. It is precisely the consideration of such intrinsically motivated *positive* reciprocity which would matter as a means to compel the trustor to trust in the first place.

In the reciprocity model presented above, norm-compliance was modeled in terms of a utility-enhancing process, displaying the intrinsic motivation and value that can stem from compliance to an internalized norm. But the internalization of a social norm during socialization is also accompanied by the installation of a "conscience"—an internal sanctioning mechanism that regulates behavior and ensures the structuring power of norms effectively by making compliance and adherence to the norm internally sanctionable (Elster 1989).[22] As it is, the moral demands on reciprocal trustworthiness that a trustee earns by honoring trust are "balanced" by the fact that the trustor holds sanctionable expectations in the form of a moral obligation to reciprocate his trust in the first place. If these moral obligations are not met, their violation triggers feelings of guilt and shame in the trustee, and anger in the trustor (Elster 2005).

A failure of trust creates psychological costs for both parties, and they increase in the degree to which a relevant norm has been internalized (Ripperger 1998: 152). The sources of guilt are diverse—fairness and equity considerations, perceived moral obligations of reciprocal response, given promises and commitments, expectations of appropriate role performance and the like can all become a matter of disappointment for the actors involved. In economic terms, "a guilt-averse player suffers from guilt to the extent he believes he hurts others relative to what they believe they will get" (Charness & Dufwenberg 2006: 1583). In contrast to purely outcome-based models of distributional fairness, guilt-aversion models are an example of psychological games in which higher-order beliefs, that is, conjectures about the other's state of mind, are critically important. With respect to the trust problem, the trustworthy or untrustworthy response by the trustee is the causal factor that triggers guilt. Essentially, if trust is failed even when the trustee believes that the trustor expects a favorable response, this induces a feeling of shame and guilt in the guilt-averse trustee, which is expressed as a disutility.

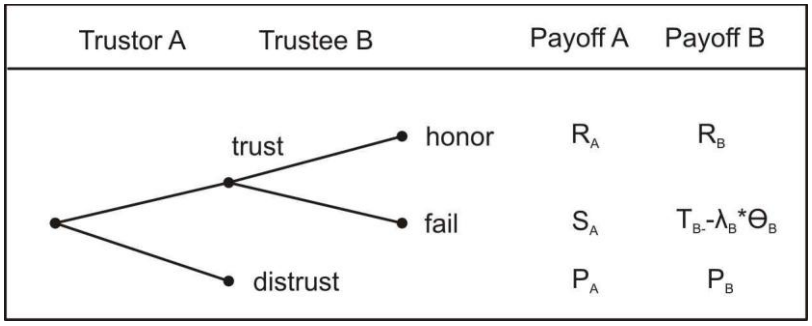
The choice of a trusting act not only indicates that the trustor rationally expects a trustworthy response, but it potentially conveys a positive appraisal of the trustee's characteristics, an acknowledgement of his abilities and integrity, and an appeal to his sense of moral duty and

[22] With respect to the sanctioning potential of norms, Elster notes: "Social norms have a grip on the mind that is due to strong emotions their violations can trigger. I believe that the emotive aspect of norms is a more fundamental feature than the more frequently cited cognitive aspects. If norms can coordinate expectations, it is only because the violation of norms is known to trigger strong negative emotions, in the violator himself and in other people" (Elster 1989: 100).

loyalty. In essence, it is the trustee's self-image to which those positive expectations cater that is "at stake" for the trustee when deciding about trustworthiness. The trustee's readiness to answer these implicit appeals with a trustworthy response is also termed *trust responsiveness* (Guerra & Zizzo 2004, Bacharach et al. 2007), that is, the "tendency to fulfill trust because you believe it has been placed in you" (Bacharach et al. 2007: 350).

An internal guilt and sanctioning mechanism can be formalized with the help of psychological games (see Battigalli & Dufwenberg 2007, 2009). When applied to the trust game, the preferences of the trustee are modeled in such a way that his utility not only depends on material payoffs, but also on his beliefs about the trustor's state of mind. If the trustee believes that honoring trust is favorably expected by the trustor, a failure of trust will lead to a decrease in utility. In a simple case, the trustee's utility function for a failure of trust is given by $U_B = T_B - \lambda_B * \Theta_B$, where Θ_B is player B's second-order belief about the trustor's first-order expectation of trustworthiness, and λ_B describes an individual parameter of "guilt sensitivity." This sensitivity parameter reflects individual learning and socialization histories and absorbs inter-individual differences in guilt-aversion aversion (figure 14):

Figure 14: Trust game with guilt aversion



The extent to which the trustee experiences guilt depends on his assessments of the trustor's state of mind: the more the trustee is convinced that trustworthy actions are actually expected by the trustor, the higher are the psychological costs that result from a failure of trust. A guilt-averse trustee will be trustworthy whenever $\Theta_B > (T_B-R_B)/\lambda_B$. This threshold decreases with guilt sensitivity and increases with the opportunity cost of not failing trust. When the trustee observes the choice of a trusting act, he updates his belief to $\Theta_B \geq (P_A-S_A) / (R_A-S_A)$. In other words, a trustee can rationally infer that the trustor's expectation must exceed the threshold derived for the simple trust game with incomplete information. This means that, since beliefs are updated during play, the trustor can use the choice of a trusting act as a strategic signal to evoke trustworthy responses, given that his belief p about Θ_B is high enough. This is known as *psychological forward induction* (Battigalli & Dufwenberg 2009). Essentially, it can be shown that models of guilt-aversion involve equilibria in which trustors rationally choose to trust in anonymous one-shot situations, given that both the number of "socialized" players in

the population and the probability of meeting a guilt-averse player are sufficiently enough (Kandel & Lazear 1992, Servatka et al. 2008).

Models of guilt-aversion can explain trustworthy responses in one-shot situations and serve as a means to encapsulate the interests of the trustor. However, empirical support for them seems to be limited. While some experimenters report results that support guilt-aversion as a motivational factor in rational choice considerations (Charness & Dufwenberg 2006, Bacharach et al. 2007, Charness & Dufwenberg 2007, Reuben et al. 2009), others report only weak to no evidence (Vanberg 2008, Ellingsen et al. 2010) and argue that it empirically plays only a minor role. An important theoretical drawback of guilt models is that they explain positive strong reciprocity, but they are unable to account for the large range of punishment activities found in human behavior (Fehr & Schmidt 2006). As with reciprocity, guilt-aversion is limited to certain domains of human action. Put sharply, when taken together, the empirical evidence seems to suggest that social preference models do sometimes play an important role during choice, but sometimes they do not.

This confronts us with a serious problem that emerges with the continuing advancement of psychological game theory and refined game-theoretic models: which model is valid, and under which circumstances? When can we assume social preferences to be part of an actor's utility function, and when not? The proposed models can explain a variety of empirical observations, but they are often limited to specific situations, sometimes to very particular experimental games; irregularities in other domains of application are routine. As the number of proposed psychological mechanisms increases and as the range of potential preferences becomes increasingly heterogeneous, the question of "which model will provide a valid explanation under which condition" becomes increasingly important—and to date, it has remained unanswered (Fehr & Schmidt 2006, Kroneberg 2006a). More pointedly, social preferences explain behavior for an exogenously given utility function, but they are not concerned with the question of its emergence. A successful definition of the situation is an implicit *ex ante* assumption that is not further problematized in economic accounts. The introduction of exogenous changes to preferences brings with it the potential charge of being an immunizing stratagem—any behavioral change can *ex post* be "explained" by referring to changed preferences (Smelser 1992, James 2002b). All in all, models of social preferences present an important development within economic theory to account for the fact that the self-interest hypothesis, as formulated in the core axioms of traditional rational choice theory, is regularly violated. But, as Fehr and Schmidt rightfully conclude: "While the current models clearly present progress relative to the self-interest approach, the evidence ... also makes it clear that further theoretical progress is warranted. There is still ample opportunity for improving our understanding of other-regarding behavior" (2006: 684).

3.3.6. The Limits of Rational Choice

The criticism of social preference models is one example of a substantial debate that has sparked around the rational choice framework, questioning its applicability and appropriateness in the context of human decision making in general. In short, the axiomatic assumptions of rational choice have been repeatedly called into question both on theoretical and empirical grounds, and researchers have consequentially started to demarcate the limits of rational choice (Elster 1979, 1986b, Hogarth & Reder 1987, Cook & Levi 1990, Coleman & Fararo 1992).[23]

Empirically, a large number of anomalies and paradoxes have been detected which are at odds with the fundamental postulates of the rationalist paradigm. For instance, the preference transitivity axiom is regularly violated by humans, even in simple choice problems, indicating that people do not conform to the principle of maximizing expected utility (Allais 1953, 1979). Decision-makers tend to give a higher weight to "known" probabilities than to "unknown" ones, and humans do not deal with ambiguity in the same way as with risk, suggesting that probabilities are not linear and additive as proposed in the SEU framework (Ellsberg 1961). What is more, risk preferences seem to differ between the domains of gain and loss; humans may be risk-seeking with respect to gains and risk-averse with respect to losses and, in addition to such loss aversion, exhibit *status quo* biases and regret (Kahneman & Tversky 1979). Biased perceptions of probabilities, that is, overestimation or underestimation of desired or undesired outcomes, can even lead to a complete reversal of preferences (Lichtenstein & Slovic 1971, Tversky et al. 1990, Slovic 1995). Researchers have also cast doubt on the assumption that decision making is independent of irrelevant situational features ("framing effects," see chapter 4.2), and have questioned the implicit assumption of reference independent utility (Tversky & Kahneman 1981, Kahneman 2003).

Quite generally, it has been found that in a majority of cases the equilibrium predictions of economic theory are reached only in very competitive environments, whereas they fail in nonmarket situations with a less ordered structure (Ostrom 2003). Psychologists have uncovered a myriad of "biases" in judgment and decision making which indicate that the normative framework of rational choice is not a valid empirical description of human decision making, and these findings have resulted in calls for the revision of the standard economic model (Mellers et al. 1998, cf. Gintis 2007). An ever-growing body of research indicates significant and systematic empirical deviations from the hypotheses of SEU theory and the standard rational choice behavioral assumptions—on the whole, "psychologists, sociologists, and economists have produced huge number of observations which cannot easily be explained within

[23] This section presents only a very brief sketch of the voluminous body of critique; it is necessarily incomplete and selective. The reader is referred to the literature presented here for a more exhaustive discussion.

the rational choice paradigm" (Boudon 2003: 8). This discrepancy and inconsistency between theoretical predictions and empirical data has challenged the postulates of rational choice theory. Most critique is directed towards the unrealistic behavioral assumptions underlying rational choice, drawing upon the fact that observed behavior is often "radically inconsistent" (Simon 1978) with the SEU framework.

For example, the assumption of full and perfect information about the environment and common knowledge of preferences and the game structure (a requirement for equilibrium prediction), has been criticized as an unrealistic idiosyncrasy of *homo oeconomicus*. Rarely are real-world actors perfectly informed about the true preferences of other players, nor are they able to immediately comprehend the complete incentive structure or all interdependencies of a decision problem. Full information assumptions have been accepted as empirically falsified even by economists. According to Gintis, rational expectations and beliefs, which represent such "always" limited information, are the most "underdeveloped member" (2007: 15) of the rational choice trilogy of expectations, preferences, and constraints because there is no compelling analytical theory of how rational agents acquire and update their beliefs. The empirical invalidity and inadequacy of the full information assumption is a serious objection to the standard model, as it cannot deliver an explanation of optimal information search and the ways by which limited information is coped with to reach equilibrium (Arrow 1987).

A closely related source of criticism is the treatment of preferences within the rational choice paradigm (March 1978, Slovic 1995, Rabin 1998, Fehr & Hoff 2011). Preferences are treated as exogenously given, and they are assumed to be stable and time-consistent for the purpose of the analysis—the qualification of stable and parsimonious preference functions ensures that the framework is not empirically empty (Vanberg 2004). Considering the available evidence, these assumptions are optimistic at best. Apart from the above-mentioned choice-inconsistencies questioning the von Neumann and Morgenstern (1944) axioms on utility functions, humans elicit preferences that change with varying time-horizons ("hyperbolic discounting"), so that it is impossible to rationally assess the best outcome in terms of utility at the instant of decision making. That is, utility functions at different points of time are intertemporally inconsistent with one another (Frederick et al. 2002). Moreover, empirical evidence suggests that preferences are often generated "on the spot" in response to the choice or judgment task, and do not enter as an invariable constant (Payne et al. 1992, Tversky & Simonson 1993). The introduction of models of social preferences in the last section illustrates the fact that researchers have started to hypothesize about different utility functions in an attempt to make economic models of decision making more realistic, and to accommodate for the variety observed in behavior and choice. However, this approach involves the methodological drawback that "rational choice theorists are forced to create new 'utility functions' for each deviation from rationality" (Weber et al. 2004: 284). These "wide" conceptions of ra-

tional choice (as opposed to "narrow" conceptions which center on pure self-interest, see Opp 1999), provide no limitations on the set of explanatory factors that might be introduced as additional utility terms, but they are "formulated so expansively that they absorb every alternative hypothesis" (Green & Shapiro 1994: 184). The scope and power of rational choice theory is severely limited by the fact that the rationalist paradigm does not offer a theory of preference formation, even when social outcomes clearly depend on preferences (Friedman & Hechter 1988). In short, while rational choice theories explain behavior for a given utility function, they cannot account for their emergence (Bicchieri 2006).

Another caveat against the orthodox rational choice paradigm concerns the assumption of perfect rationality and strict utility maximization. *Homo oeconomicus* is often envisioned as being a "hyper-rational" (Weber et al. 2004) agent, a sort of walking computer that, given preferences and beliefs, can instantaneously calculate the costs and benefits of his actions. As Selten puts it, "full rationality requires unlimited cognitive abilities. Fully rational man is a mythical hero who knows the solutions to all mathematical problems and can immediately perform all computations, regardless of how difficult they are" (Selten 2001: 14). In contrast, psychological research in human judgment and decision making has convincingly demonstrated that cognitive capacities for rational calculations of the sort proposed in economic models are limited (Payne et al. 1992, Gigerenzer & Selten 2001). This indicates that the decision-making behavior of "real" human beings cannot conform to the ideal of full rationality. The postulation of perfect rationality has been supplemented by propositions of human *bounded rationality*, as introduced, for example, in the works of Simon (1955, 1978).[24] According to Simon, human decision making can be more adequately described as a process of *satisficing*, in contrast to the ideal-type *maximizing* behavior of the perfectly rational actor. A satisficing actor will be "doing the best" under the irremovable constraints of limited cognitive capacity and notoriously scarce and uncertain information, and will process information only to the extent that suffices to reach a certain aspiration-level of utility. While the axioms of rational choice postulate a form of "substantial rationality," it is preferable to regard the "procedural rationality" of the decision-making process as well.[25] When formalizing accounts of bounded rationality, the cost of information processing and search have to be included into the decision-making process (Riker & Ordeshook 1973, Heiner 1983). In sum, perfect rationality is a rather hypothetical "limiting case" (Ostrom 2003) of human bounded rationality.

[24] "Rationality is bounded when it falls short of omniscience. And the failures of omniscience are largely failures of knowing all the alternatives, uncertainty about relevant exogenous states, and inability to calculate consequences" (Simon 1979: 502); see also the very informative discussion of the concept of bounded rationality in Selten (2001).

[25] This notion of adaptive bounded rationality is consistent with the theoretical frameworks developed in evolutionary psychology (Tooby & Cosmides 1992), evolutionary game theory and biology (Gintis 2000), and social cognition (Chaiken & Trope 1999), which will be discussed in chapter 4.

Furthermore, in the rational choice paradigm rationality ultimately centers on intentionality, self-interest, maximization, and the consequentialism of action and choice. Even in extended models of "wide" rational choice accounts, actions are always instrumental and outcome-oriented. In contrast, sociological theorists throughout have stressed the noninstrumental character of action, as exemplified, for example, by Weber's distinction of "axiological" and "instrumental" rationality (Boudon 2003). In Weber's sense, actions are always meaningful and should be understood as based on reason, but such reasons can take forms other than cost-benefit considerations. The notion of rationality can therefore be noninstrumental, and must be given a new "content" in some situations. According to Elster, the instrumentality of actions cannot be extended to domains such as friendship, love, and respect: "Altruism, trust and solidarity are genuine social phenomena and cannot be dissolved into ultra-subtle forms of self-interest" (1979: 146). He contends that the rational choice paradigm cannot account for these aspects of social life and turns to norms as an "autonomous" motivational factor which lies outside of self-interested utility considerations. In essence, norms "override" the rationality of self-interest. In a similar fashion, researchers have stressed the importance of rules and adaptive rule-based choices as an *alternative* to consequence-based maximization (March & Olsen 1989, March 1994, Vanberg 1994, 2002, 2004). This complements the conceptions of rule-based trust, based on "shallow" decision making. But as we have seen, norms, rules, and social institutions do not stand outside the utility considerations of rational actors in the rationalist paradigm. They matter only insofar as they are parameters in the actor's calculation of whether or not to choose some course of action. It is, however, questionable (both on theoretical and empirical grounds) whether strict rationality and self-interest maximization constitute the sole and only "logic" of decision making in social settings (Messick 1999, cf. Kirchgässner 2008).

Taken together, a number of theoretical arguments demarcate the limits of rational choice as a general framework of decision making, adding to the enormous body of empirical evidence that highlights violations of its fundamental postulates. These challenges naturally transfer to the problem of modeling and explaining interpersonal trust with its help. In fact, they are of particular importance when studying interpersonal trust, where a tension between "rational" and "irrational," conditional and unconditional, cognitive-based and affect-based conceptualizations has always been a fundamental aspect of theorizing. With respect to the game-theoretic approaches, Hardin aptly notes that, "despite their clarity in many respects, the game representations of various interactions do not unambiguously tell us about trust. The broad range of potential reasons for player's choices is not narrowed to trust, self-interest, normative commitments or any other motivation" (Hardin 2003: 98).

3.4. Is Trust Rational?

The explanative strategy adopted by economists in the explanation of interpersonal trust is to "create" and model the incentives for the trustee to be trustworthy. Trustees honor or fail trust because of the costs and benefits attached to honoring or breaking trust. By making the trustworthy option more desirable in terms of individual utility, whether internally or externally motivated, one can thereby "encapsulate" the interests of the trustor in the trustee's rationale. The choice of a trusting act is regarded as a conscious, maximizing and deliberate decision based on the trustor's expectation of trustworthiness (Coleman 1990, Hardin 2001). The cost-benefit considerations may take into account contextual factors and incentives from social embeddedness (dyadic, network, and institutional learning and control), they may include conjectures about trustee characteristics (benevolence, competence, integrity, and predictability) and his state of mind (internalization of reciprocity norm, guilt-aversion), they may be "biased" by generalized expectations and individual predispositions to trust, and we can justly assume that they are backed-up by the prevalent culture of trust inherent in the social system and the structural assurance and situational normality beliefs conferred thereby—even when these aspects are not part of explicit economic modeling. The higher a trustor's expectation in a particular situation, the more likely it is that he decides to choose a trusting act. As it is, the rational choice perspective of trust takes a very specific perspective on the phenomenological foundation characterizing interpersonal trust: it is a deliberate and "effortful" decision that takes place after an assessment of all relevant incentives which might potentially motivate the trustee. As Möllering is apt to point out, the conceptual starting point for economic accounts is "first and foremost wariness, if not paranoia, of opportunism ... the underlying models are conservative in that they emphasize the pervasiveness of opportunism, the risk of exploitation and the costs of safeguards against the detrimental actions of others" (2006b: 24).

As we have argued in chapter 2, assumptions concerning the subjective experience of trust are most divergent. It is a matter of lively debate whether trust, as a mental phenomenon, is connected to reflective processes and the consultation of reason; whether actors do trust "consciously" and calculate the risks and utilities involved and act upon them—or whether trust includes the suspension of doubt, a perception the trust problem "as if" there were none, and therefore a considerable unawareness of the risks and incentives. The tension between these conflicting theoretical perspectives and the promise of their future reconciliation has been a prime motivation for this work. The exposition of the game-theoretic perspective in the last sections has presented us with a perspective on trust in which the role of reason is clear-cut: trustor and trustor do explicitly reason about their choices, taking into account their vulnerabilities and all relevant risks and incentives. The basal logic of the choice of a trusting act is similar to a "bet" on risky alternatives, and the rational actor can make such a choice because he compares and evaluates the alternatives at hand. This necessitates the stability, unambigui-

ty, and cognitive availability of expectations. The choice of a trusting act is then a behavioral "by-product" of these expectations. All extended models of the basic trust game, such as those modeling embeddedness and social preferences, or adding additional incentives for trust and trustworthiness, conform to the same principle. The cognitive-reflective perspective on trust is vividly expressed in Luhmann's conceptual distinction between familiarity, confidence, and trust: "The distinction between confidence and trust depends on perception and attribution. If you don't consider alternatives (every morning you leave the house without a weapon!), you are in a situation of confidence. If you *choose* one action in preference to others in spite of the possibility of being disappointed by the actions of others, you *define* the situation as one of trust" (Luhmann 1988: 97, emphasis added). In this cognitive perspective, trust always includes a conscious acceptance of the risk and vulnerabilities included in the trusting act; it is conceived of as a rational choice among risky alternatives.

In contrast to economic accounts, many psychological and sociological conceptualizations emphasize the prereflective nature, emotionality, and unconditionality of trust. From that perspective, trust "can also be shown to be thoughtless, careless and routinized, and thus to require no unnecessary expenditure of consciousness, especially if expectation approaches certainty" (Luhmann 1979: 24). Luhmann's claim is reminiscent of the idea of a noncognitive leap of faith (Lewis & Weigert 1985b) enabled by trust, a suspension of risk built on taken-for-grantedness (Möllering 2006b), and it also hints to the subjective experience of trust as a predominantly emotional attitude (Holmes 1991, Lahno 2001, Karen 1996), to mutual identification and a "non-cognitive security" (Becker 1996) in which the trustor's "trustful" state of mind becomes the interpretive lens that interferes with a cold cognitive assessment of trustworthiness. As Karen puts it, "the harms they [the trustees] might cause through failure of goodwill are not in view because the possibility that their will is other than good is not in view" (1996: 12). While uncertainty is seen as part of the objective structure, it is not regarded as part of the subjective perception. Essentially, trust is intangible to reflective reasoning because it logically precedes deliberation; it is already part of the process of definition of the situation. This alternative viewpoint must be taken into account in a broad conceptualization of trust.

To some researchers, the economic approach also poses a serious definitional problem because trust, based on prudence, is not considered to be proper trust at all (March & Olsen 1989, Williamson 1993, Miller 2001, James 2002b).[26] In this line, Williamson laments that

[26] Such definitional quarrels are nonetheless irrelevant in the context of this work. Neither are "narrow" definitions of trust helpful when elaborating on a broad conceptualization of interpersonal trust, nor it is the researcher's task to speculate about the "essential" and "realistic" nature of trust—as conceptual nominalists, we have to acknowledge the fact that trust can have a broad phenomenal and experiential basis, and therefore need to include, not rule out, "rational" forms of trust in our theoretical framework. In fact, the related arguments beg the question of why people trust when doing so does not involve incentives for the trustee.

"calculative trust is a contradiction in terms" (1993: 463). Essentially, economic approaches are "designed to explain trust away" (Möllering 2006b: 43) because the incentive structure, when it is appropriate for a trustworthy response, does remove the vulnerability to exploitation that gives trust its very meaning (Miller 2001, James 2002b). As Lewis and Weigert note, "trust begins where prediction ends. The overrationalized conception of trust, by reducing it to a conscious, cognitive state presumably evidenced by cooperative behavior, totally ignores the emotional nature of trust" (1985b: 976). In fact, even in the social preference models introduced above, norms and emotions remain an instrumental means to achieve a desired end. In criticism of Coleman's approach, Misztal laments that "self-interest exploits social norms to punish untrustworthiness" (1996: 80). These shortcomings of economic accounts, even more so in light of the irrefutable limitations of the rational choice paradigm, have stirred concern even among rational choice advocates of trust, putting its descriptive adequacy into question (Kramer 1999).

Hardin, arguing in favor of the rationalist perspective, rightly admits: "Trust is not a risk or a gamble. It is, of course, risky to put myself in a position to be harmed or benefited by another. But I do not calculate the risk and then additionally decide to trust you; my estimation of the risk is my degree of trust in you. Again, I do not typically choose to trust and therefore act. Rather, I do trust and therefore choose to act" (Hardin 1993: 516). But such an argument completely fails to explain trust in the first place. The way out taken by him and most economists (e.g. Coleman 1990) is to *equate* trust and trusting expectations—trusting expectations are regarded causal antecedents to behavioral trust. Consequentially, there is nothing beyond the cognitive categories of trust-related knowledge (as expressed in the trustor's expectation) which enable the rational choice of a trusting act. Thus, the implicit assumption in rational accounts is that "the lands of interpretation and expectation are directly connected (if not one and the same)," as Möllering (2001: 413) points out.

Yet, numerous trust researchers contend that expectations—if perceived at all—are merely a consequence of prereflective processes in which trust already plays a role. In this perspective, trust is not reducible to expectations, but it constitutes a logical antecedent. This view is rightly expressed in the idea of a "cognitive leap" and suspension as a consequence of interpretation, and it is implicitly made in the distinction between conditional and unconditional forms of trust (Jones & George 1998), the many claims of trust as "an unintended outcome of *routine* social life" (Misztal 2001: 323, emphasis added), based on taken-for-grantedness and situational normality. In the words of Elster, cultural tradition and social norms enable the choice of a trusting act by "overriding" rational considerations. These forms of shallow trust are not based on effortful decision-making processes, but indicate a rather low level of rationality and cognitive deliberation (Messick & Kramer 2001). In a sense, they describe the choice of a trusting act as based on simple heuristic processes, substituting the idea of rational

choice with a logic of appropriateness in which the adaptive use of rules, roles, routines and emotions as rules of thumb, under the constraints of bounded rationality, helps to "quick-step" interpersonal trust.

The preceding discussion suggests that trust and rationality are closely intertwined. Whether trust is a rational choice or not is a key question to which trust researchers within the different research paradigms have given multifaceted and divergent answers. While economic approaches emphasize rationality, many sociological and psychological accounts maintain that trust indicates the absence of rationality. If we take a look at the broad picture that arises, one conclusion we might draw is that the approaches are inconsistent, contradictive, disconnected, and stand next to each other in a relatively independent fashion. At first sight, any further theoretical integration is prevented because of the huge differences in the underlying assumptions concerning interpretation, choice, and the degree of rationality involved.

A more fruitful alternative is to ask for the common ground that would allow "rational" and "nonrational" accounts of trust to be united and integrated. A most promising avenue in this respect is to put the degree of rationality involved in interpretation and choice into the focus of trust research. The reason for this stipulation is simple: considering all, rationality seems to be a key dimension that helps us to discriminate and integrate the various typologies of trust which have been proposed. Cognition-based versus affect-based, calculus-based versus identification-based, conditional versus unconditional trust: most typologies implicitly rest on specific assumptions concerning the amount of rationality involved in the choice of a trusting act. Moreover, they differentiate trust with respect to the categories of trust-related knowledge that are applied.

The two aspects (category of trust-related knowledge and the degree of rationality involved) are often mixed up and woven together in an inconsistent and contradictory way. In essence, rationality is not regarded as a proper and independent dimension of the typological space of trust, and thus is not "orthogonal" to the categories of knowledge used to solve a trust problem. Instead, types of knowledge and their mode of application are portrayed as fixed and interwoven. But one apparent reason for the diversity of typologies proposed is the fact that knowledge can be processed and applied in different ways, and whether we focus on more automatic or rational processes in interpretation and choice will eventually necessitate creating a new "type" for each category of trust-related knowledge and each degree of rationality we assume.

The solution that will be put forth in the following is to regard rationality as an endogenous result of internal cognitive processes. In this sense, it has to be understood as a fundamental ingredient in the solution of trust problems, and a basic dimension of the trust concept itself. This necessitates a turn to social-psychological frameworks, which have started to accumulate

evidence of human *adaptive rationality*. In the "dual-processing" approach to cognition, rationality is seen as bounded but also highly flexible and adaptive to internal and external constraints. The degree of rationality involved in interpretation and choice thus will dynamically change, and it is principally independent of the categories of trust-related knowledge that are applied. Ultimately, adopting such a perspective of adaptive rationality paves the way to a broad integration of different approaches to trust under a common theoretical umbrella. In order to move towards such broad account, we will now turn to the concept of adaptive rationality.

4. Trust and Adaptive Rationality

> "Human rational behavior is shaped by a scissors whose two blades are the structure of task environments and the computational abilities of the actor" (Simon 1990: 7).

In recent years, an ever-growing part of trust research has been concerned with the various ways in which cognitive processes influence, bias, and determine the choice of a trusting act. An idea that has gained increasing attention is that there in fact exist a number of different "routes to trust" which can be taken by a trustor faced with a trust problem. The propositions are very diverse, but they have in common the principal idea that trust may be the product of cognitive shortcuts which help to facilitate information processing and reduce the cognitive load of the trustor, in order to free up sparse cognitive resources and processing capacity. Essentially, they are indicative of a variable degree of rationality involved in the choice of a trusting act and, more generally, of the *adaptive rationality* inherent in interpretation and choice.

To grasp the concept of adaptive rationality, we will take a close look at the "dual process paradigm," which explicitly takes into account the particularities of the human cognitive system with respect to individual-level rationality. The concepts developed in this research tradition can fruitfully inform trust research. Broadly speaking, dual-process models assume that human cognition may occur in either a rational or an automatic mode. These are not only characterized by very distinct functional properties, but in fact make use of different neuronal systems in the brain, and are thus "hard-wired" into the human cognitive architecture. While the automatic mode is intuitive, emotional, fast, effortless, and associative, the rational mode is slow, serial, effortful, and controlled. This dual-processing framework can be usefully applied to trust problems, as it suggests that different processing modes—that is, different degrees of rationality—are involved in interpretation and choice when solving a trust problem.

The contingent and flexible use of different processing strategies in trust problems also points to the context-dependence of information processing, and, incidentally, to the context-dependence of trust. Exploring the role of the context in trust research, we find that its influence is most often taken for granted and left implicit. However, it is a key factor in dual-processing accounts of cognition, governing the mode of information processing that individuals adopt. It influences the accessibility and activation of knowledge, and determines, among other factors, whether and when trustors use more "heuristic" or more "elaborate" strategies to solve a trust problem. In short, when thinking trust in terms of adaptive rationality, we must respect and account for the context in our theoretical models.

In a logical next step, we need to ask about the determinants of information processing and answer the pressing question of how and when the degree of rationality changes in response to

internal and external factors. The dual-processing paradigm has carved out four central determinants of the processing mode: (1) opportunities to engage in controlled processing; (2) the motivation to do so; (3) the availability, accessibility, and context-dependent applicability of knowledge; and (4) tradeoffs between effort and accuracy. However, as the dual-processing paradigm is scattered among different thematic domains and research traditions, specific accounts rarely inform each other or collect existing knowledge into a broad picture combining these determinants in a general framework. Of most immediate concern is the fact that there is no theoretical model available that tells us precisely how the four determinants are functionally related to each other, and how adaptive rationality may, in fact, be modeled. Most theories are content with listing a set of "moderators," leaving the question of their precise interplay open. Moreover, existing theory does not explain how interpretation and choice are connected, that is, how we could causally model the links between adaptive rationality, interpretation, and choice.

To this end, I will discuss the "Model of Frame Selection" (MFS), a sociological model of adaptive rationality that has been developed over the last two decades with the aim of providing the general theory of action. A unique feature of this theory is that it connects adaptive rationality to interpretation and choice, by focusing on the process of "mode selection." The process of mode selection is modeled and conceived of as an autonomous, regulative achievement of the cognitive system, in which the degree of rationality that actors use during knowledge application is determined. The model derives a clear and formally precise formulation of the process of mode selection from minimal assumptions; it also spells out the "selection rules" which govern the activation of mental schemata in the different processing modes, and thus establishes a causal link between an actor's adaptive rationality, interpretation, and choice.

In developing this perspective of adaptive rationality, I use the model to causally explain both conditional and unconditional types of trust. With a tractable behavioral model at hand, it is apparent that rationality must be regarded as a key dimension of the trust concept. Importantly, the degree of rationality involved in interpretation and choice can dynamically change. It is not fixed, and therefore either automatic *or* rational processes may prevail both during the definition of the situation, and the choice of a trusting act. The perspective of adaptive rationality helps to integrate seemingly contradictory accounts of trust, because it gives evidence for the fact that trust researchers focus on different aspects (interpretation *or* choice) of the trust problem, and assume different modes (rational *or* automatic processing) when theorizing about a particular solution. The chapter closes with the development of a theoretical model that pinpoints the mode selection thresholds governing interpretation and choice in a trust problem, and with a definition of trust that takes care of actors' adaptive rationality and with a state-

ment of general model propositions. The model will be used in chapter 6 to develop and test empirical hypotheses in a controlled laboratory experiment.

4.1. Different Routes to Trust

While the mainstream of trust research draws it foundations from the rationalist paradigm, a number of contributions—backed up by advancements in cognitive psychology—suggest that trustors often use "mental shortcuts" and heuristic strategies to solve a trust problem. These approaches not only accept human bounded rationality as a foundational starting point for theorizing, but they also demonstrate the implications of the limited cognitive capacity and bounded rationality of humans when thinking about interpersonal trust. The following paragraphs present a general overview of the variety of models which have been proposed in favor of a bounded-rationality approach to trust.

To begin with, some authors hypothesize specific heuristics which work as a direct shortcut to generate interpersonal trust. For example, Burnham et al. (2000) postulate the existence of a mental "friend-or-foe" module (FOF) and its use in trust problems. The FOF module is conceptualized to be an adaptive mechanism of the brain which automatically changes perceived cooperativeness or competitiveness, releasing trustors from the need for otherwise costly and effortful mental accounting. "Friend-or-Foe detection primes you for greater expected benefits than without it. It sets you up *preconsciously* for a maximizing decision. If you get surprised, you have to reconsider with more cognitive resources ... It is a *heuristic routine* which saves having to think carefully about every decision, but it is not an irreversible commitment" (Burnham et al. 2000: 61, emphasis added). The FOF module is triggered by immediate situational cues, and it is considered to be "part of the human autonomic decision processing capacity" (ibid.).

Generally speaking, FOF detection alters the perceived likelihood of a trustworthy response before trustors consciously perceive it. In other words, it directly influences the formation of the trustor's expectation of trustworthiness. In the case of a "foe" being detected, expectations become unfavorable, because the cognitive system becomes attentive and suspicious to opportunism and breaches of trust. Unfortunately, Burnham et al. (2000) do not further specify or model the FOF mechanism. Empirically, they demonstrate that a change in the experimental instructions of a repeated trust game (referring to "partners" versus "opponents") has dramatic effects on the observed levels of trust and trustworthiness over an extended period of time.

In the same line, Yamagishi et al. (2007) hypothesize the "social exchange heuristic" (SEH), which—if activated—completely suppresses the perception of opportunities for defection. The heuristic is also regarded as an evolutionary adaptation of the human organism for facilitating social exchange, and is a "cognitive bias that perceives free riding in a situation as nei-

ther possible nor desirable" (ibid. 10). The SEH is activated by cues that hint at the presence of a situation of social exchange. More concretely, actors are assumed to make subjective inferences about the true state of the world, and evaluate the potential errors of this inference process. According to the authors, the inference evaluation process is unconscious and automatic; it is concerned with the question of whether or not free riding is likely to be detected, and whether punishment is a credible threat. The inference process has two possible outcomes which correspond to the possible states of the world ("sanctioning" or "not sanctioning" free riders). Accordingly, two different errors may be committed, if the result of the inference process does not correspond to the true (Figure 15):

Figure 15: Inferences and the SEH, adapted from Yamagishi et al. (2007: 266)

		True State of the World	
		No Detection / Sanctioning	Detection / Sanctioning
Inference	No Detection / or Sanctioning	Correct Inference Outcome: Saving in cost of cooperation	Type II Error Outcome: Punishment for detected free-riding
	Detection / Sanctioning	Type I Error Outcome: No Saving in costs of cooperation	Correct Inference Outcome: Gains from mutual Cooperation

If an actor makes the "sanctioning" inference (that is, if the presence of sanctioning mechanisms is detected as a credible threat), this automatically activates the SEH. Thus, whenever a situation is defined as being under social control, the alternative of defection is simply suppressed from perception. Applied to a trust problem, both trustor and trustee could define the situation as one of social exchange and, if so, they would not even consider the possibility for a failure of trust. This clearly contradicts standard economic models, where the prospective costs of defection are weighed against the prospective gains. The SEH is conceptualized as an adaptive-evolutionary heuristic for the detection of situations of social exchange, and, similar to the FOF module, is assumed to be "hard wired" into the human brain.

The models of Burnham et al. (2000) and Yamagishi et al. (2007) are formulated in the tradition of evolutionary cognitive psychology (Cosmides & Tooby 1992). This paradigm regards heuristics as specialized adaptations of the mind to solve particular problems. As Cosmides and Tooby argue, social cognition consists of a rich set of "dedicated, functionally specialized, interrelated modules to collectively guide thought and behavior" (1992: 163). Furthermore, cognitive adaptations are highly domain-specific. They argue that the domain of social exchange has been of particular importance in human evolutionary history, which is why spe-

cific adaptations for social exchange ("social heuristics") are likely to have evolved over time. This heuristics perspective can be contrasted to the "general-purpose reasoning" approach prevalent in the social sciences, where human problem-solving is assumed to be achieved by content-independent procedures, such as logical inference and propositional calculus. As Cosmides and Tooby argue, the domain of social exchange activates behavioral and inferential rules that cannot be accounted for in terms of general-purpose reasoning.

Although the models introduced above emphasize human bounded rationality, the fact that they each propose a very fixed and specific heuristic to solve the trust problem is limiting to a broad conceptualization of trust. Obviously, invoking a specific heuristic can explain interpersonal trust in some instances, but other avenues to the choice of a trusting act can be imagined and should not be omitted. Following this approach, we would have to add to the list, for example, a "feeling-as-information" heuristic to account for affect-based types of interpersonal trust; a "relational schema" heuristic to explain unconditional identification-based trust; and a "routine-application-of-rules" heuristic to account for instances of rule-based institutional trust; and so on (for a more detailed discussion of heuristics and their use, see section 4.2.3). On the other hand, the choice of a trusting act may very well be subject to cost-benefit considerations and attempts to make rational inferences about trustworthiness. That is, the "trust as a social heuristic" approach is informative, but incomplete when it comes to specifying the broad phenomenological foundation of interpersonal trust. Nor does it capture the full range of phenomena which we can relate to interpretation and choice in trust problems.

One possible solution to this problem of "incompleteness" is to ask when and how the different heuristics govern interpretation and choice, and when they are not used but instead replaced by a more elaborate reasoning and inference process. In this line, Fehr and Tyran (2008), examining the process of expectation formation in simple "price-setting games," demand that the *degree of rationality* is treated as an *endogenous* variable, which must be related to the costs of error detection, and the costs of making false decisions. They note that "a key question ... is to identify the conditions under which limited rationality occurs and when it affects aggregate outcomes in economic interactions ... The strategic environment may change the amount of individual level irrationality" (ibid. 354). Unfortunately, Fehr and Tyran do not offer a formal model of this proposition. But the idea of treating the degree of rationality as an endogenous variable has a promising theoretical advantage over the trust-as-heuristic approach in offering a general explanation of how and when the use of heuristics versus rational decision-making processes will occur. In short, it suggests thinking about actors' *adaptive rationality*, which is not just bounded, but at the same time flexible and adaptive to the environment and internal constraints.

Adaptive rationality, if it does form part of human cognitive architecture, will come to bear in trust problems as well. The idea that trustors might vary in the degree of rationality they adopt

in solving a trust problem has been proposed by Ziegler (1998), who discusses individual forecasting abilities with respect to the precision of trustworthiness expectations, and relates them to the mental costs incurred by cognitive processing and increased attention. He argues that a higher precision level of expectations is connected to higher mental costs and therefore, when limited cognitive capacities exist, such precision is often not attainable. A related argument was made by Williams (2001), who proposes that the processing of trust-related knowledge regularly happens in a heuristic, category-based fashion. The category-based processing of trust-related information is not something actors always and intentionally opt for, but a result of limited cognitive capacities. It triggers "category-based affect and beliefs," which then influence trusting expectations and intentions. Principally, both Ziegler and Williams argue that the use of heuristics is not simply "hard-wired," but is instead dependent on internal and external conditions to which the cognitive system needs to respond.

Even more thought-provokingly, Hill and O'Hara (2006) sketch a theory of trust in which both conscious and unconscious strategies for the placement of trust exist. They argue that the automatic-spontaneous choice of a trusting act is a heuristic "default rule," and they substantiate their argument by reference to the "feeling as information" paradigm (Clore 2000), according to which subjective emotional experiences can be heuristically used as an internal source of information during judgment and decision making. Likewise, Keren (2007) argues that, "while trust is an indispensable component of our daily life, it is not consistently (and continuously) raised in our awareness. Unless there is a reason, or unless primed in one way or the other, the question of trust remains in a dormant state. In most situations, and in most of our daily social encounters, as long as the assessed risk is sufficiently small, we tend to *assume trust by default*" (ibid. 252, emphasis added). In short, given that humans suffer from bounded rationality and scarce cognitive resources, the choice of a trusting act may sometimes be characterized by the involvement of a rather low level of cognitive effort; it may be warranted as a "default" decision, rooted in a state of routine springing from taken-for-grantedness; or it may be directly motivated from feelings and emotions, rather than being the result of an explicit reason-based judgment.

This suggests that flexible information-processing strategies and subjective experiences can be fruitfully combined into a broad conception of trust. In a model by Schul et al. (2008), trust and distrust span a continuum of mental states which, at the endpoints involve (1) the use of routine strategies and a subjective feeling of security in the state of trust, and (2) the nonroutine use of elaborated processing strategies, linked to a feeling of doubt, in the state of distrust. The temporary state of the mental system on this dimension of adaptive rationality determines the processing, acquisition, and elaboration of new information and its integration into judgments. As the authors argue, trust is intrinsically connected to the use of "routine" information-processing strategies and always accompanied by subjective security, while distrust

must be regarded as a result of more "elaborate" strategies that go along with a subjective state of doubt. However, their model does not explicate how and when a shift in the continuum of processing strategies occurs, and it neglects the possibility that trust may also spring from doubtful, rational inference processes.[1] Such a "dualistic" conception of trust, which explicitly includes "routine" action and "elaborate" decision-making strategies, is also empirically supported by a study of Krueger et al. (2007). Based on the analysis of fMRI data and decision times, they separate conditional trusting strategies from unconditional ones. Not only do these strategies differ in the cognitive costs associated with them, but their existence can be traced back to the preferential activation of distinct neuronal systems.

Clearly, the contributions reviewed above take a very different view of interpersonal trust than does a pure rational-choice approach. The models point to processes which precede the choice of a trusting act and structure the way in which the environment and the trust problem itself is perceived; in effect, they demonstrate potential candidates for an explanation of the "leap of faith," which, as Möllering argues, "is far from rational" (2001: 411). This review leads us to two important conclusions. First, the perspective of a preconscious, mental structuring of perception demonstrates the importance of interpretation and the definition of the situation as a substantial aspect of interpersonal trust. We have to clearly differentiate between the processes of the *definition of the situation* and the subsequent *choice of action* if we want to disentangle the phenomenon of interpersonal trust. Both aspects—interpretation and choice—will have to be respected simultaneously, though separately.

Second, these authors suggest that interpretation and the choice of a trusting act may be marked by an adaptive use of different processing strategies. Far from being self-evident, researchers often implicitly assume that the choice of a trusting act occurs with a flexible degree of rationality. It is instructive to reinspect Luhmann's apparently cognitive account under the rubric of adaptive rationality; as he contends, "trust *merges gradually* into expectations of continuity, which are formed as the firm guidelines by which to conduct our everyday lives" (Luhmann 1979: 25, emphasis added). Trust can be defined by the way in which information is dealt with in a particular situation, meaning that "primary support of trust comes from the *functions* it plays *in the ordering of information processing* internal to the system, rather than directly from guarantees originating in the environment" (ibid. 27, emphasis added). With his distinction between trust and confidence, Luhmann took a particular position concerning the rationality of trust, putting it close to the cognitive, rational-choice perspective. But as we have argued previously, the conceptual distinction between trust, familiarity, and confidence

[1] Of course this is, above all, a definitional problem. Obviously, the definition of trust as proposed by Schul et al. (2008) refers to trust as a state of perceived security, "non-reflectiveness," and the absence of doubt. In a broad conceptualization of trust, limiting the phenomenological basis in this way is unnecessary restrictive. Another problem with their approach is that adaptive rationality and trust are virtually the same—there is no difference between the prevalent processing state and the ascription of "trust" to the trustor, once the information-processing state is known.

is not well-defined and is fraught with inconsistencies on account of their gradual nature and mutual overlap (Endress 2001). All in all, the question of *how* information is processed in a trust problem seems to be a key aspect of the trust phenomenon which needs to be answered.

Apart from the necessity of analytically separating interpretation from choice, it is thus equally important to be precise about the processing strategies involved in each stage, in order to improve our understanding of interpersonal trust. The important lesson that can be taken is that, "whether or not action is founded on trust, amounts to an essential distinction in the rationality of action which appears capable of attainment" (Luhmann 1979: 25). We would have to add that, "whether or not the choice of a trusting act is conditional or unconditional amounts to an essential distinction in the degree of rationality in action." As we will see, cognitive psychology has drawn a picture of *homo sapiens* that testifies to the idea of human adaptive rationality, bounded in comparison to the ideal-type rational actor, yet flexible and highly adaptive to situational constraints, and by no means inefficient. Taken together, the discussion suggests that, apart from the missing link of interpretation and the subjective definition of the situation, a broad conceptualization of trust must respect a second factor fundamental to the emergence of interpersonal trust—namely, the *degree of rationality* involved in interpretation and the choice of a trusting act. A broad theory of trust will have to specify the process leading to its endogenous determination. In fact, it will be argued in the following that a broad conceptualization of trust must respect endogenous, *adaptive rationality as a fundamental dimension of the trust phenomenon* itself.

But when will the choice of a trusting act be guided by routine, emotions, and heuristics, and when will it approximate a rational choice? Can we say more about the degree of rationality involved in interpretation and choice? As we will see, adaptive rationality is not a black box, and in order to understand its operation and functioning, we must return to the missing link of interpretation. Both the definition of the situation and variable rationality go hand in hand, and they are of primary importance to the phenomenon of interpersonal trust.

4.2. Adaptive Rationality

4.2.1. The Dual-Process Paradigm

A theoretical paradigm that explicitly takes into account human variable rationality is the so-called "dual-process" paradigm (see Chaiken & Trope 1999, Smith & DeCoster 2000, Kahneman 2003, Evans 2008). The principal assumption that unites researchers in this tradition is that human information processing is accomplished by two underlying, yet fundamentally distinct, cognitive systems, and that humans flexibly use both "routes" and the information-processing strategies associated with their activation. Generally speaking, the models postulate the existence of two ideal-type *modes of information processing*: a parsimonious

"heuristic" mode—applied whenever the situation suggests the applicability of simple rules and associations—and an "elaborate" mode, which demands that cognitive capacities and sufficient motivation are available. While the first mode can be interpreted as a quick-and-dirty human approach for arriving at sufficiently good answers effortlessly and efficiently, the second mode involves a problem being solved by effortful mental operations, that is, by hard thinking and reasoning (Smith & DeCoster 2000). Dual-process models have been applied in many areas of research, such as social cognition (Fiske & Taylor 1991, Liberman 2006), persuasion research (Petty & Cacioppo 1986), judgment under uncertainty (Tversky & Kahneman 1983), and choice (Camerer et al. 2005). Recently, the framework has received support up from neuroscience studies linking the theoretical concepts of the framework to the underlying neuronal systems of the brain (Lieberman 2007, Rilling & Sanfey 2011).

Before we continue, it is necessary to clarify some terms which need a more detailed specification at this point. The term *perception* describes "the interface between the outer and the inner world" (Bodenhausen & Hugenberg 2009: 2). Stimuli from the environment create signals (visual, auditory, etc.) that can be sensed. Perception means that the perceiver converts these signals into the basic, psychologically meaningful representations that define his or her subjective experience of the outer world, while using the different processing routes. Of course, previously unknown, unfamiliar, and new sensory input can also be "perceived": the novelty of the sensory input is itself a meaningful perception with direct consequences for attention and higher-order inference (see below). If a meaningful percept is achieved, it can serve as an input to higher-order cognition, such as thinking, reasoning, or inference. Thus, perception is a basic process that precedes most other activities relating to the outer world. For example, the visual perception of a number of objects ("book," "Mr. Smith," "table," "child") can deliver the input for higher-order inference processes, such as a subjective definition of the situation, which is primarily concerned with interpreting complex social situations ("Mr. Smith is reading a book to his daughter at the table"). But the basic components—material objects, people, actions, symbols, social contexts—have already been perceived.

While perception can be understood as the first step in social cognition, *attention* must be regarded as the first step in perception (Bodenhausen & Hugenberg 2009). In the fashion of a "spotlight" or a "zooming lens," attention puts into focus only a limited number of stimuli at one time. A small number of stimuli from the environment receive attention, are selected for further scrutiny, and reach the threshold of awareness, while many others receive little attention—attention is a scarce resource. The question of which stimuli will be ignored and which will be attended to is one of the central problems of the cognitive system, and it is generally solved in terms of *selective attention*. This necessitates that information must be initially and preattentively processed to some extent. Preattentive scans of the environment are necessarily fast, work on a low level of stimulus features, and directly access the sensory system. These

"natural assessments" (Kahneman & Frederick 2002) of the cognitive system include reconnaissance of physical properties such as size, distance, loudness, and speed, and of more abstract properties, such as similarity, surprise, or affective valence. The capturing of attention can occur in an "active" or a "passive" fashion. It is passive (or "bottom-up") when attention is allocated automatically as a reaction to stimuli in the environment, for example because they appear quickly, surprisingly, and without warning, when they are inconsistent with stored schematic knowledge, or when they are evaluated as a threat. Likewise, unknown, novel and atypical experiences automatically attract our attention. A term that is often used to describe the fact that an event or stimulus receives selective attention is *salience* (Higgins 1996). As Higgins points out, salience "refers to something that does not occur until exposure to the stimulus, and that occurs without a prior set for a particular kind of stimulus, such as a belief about or search for a particular category" (ibid. 135). Salience is itself context-dependent: in a red world, grey is salient. On the other hand, attention can be captured actively (or "top-down") when the subjective state of the perceiver influences its allocation. For example, preexisting affective states (fear, anxiety, happiness etc.), goals, expectations, and other activated mental schemata can guide our attention, direct the activation of knowledge, and have a direct influence on how and where attention is focused (Bodenhausen & Hugenberg 2009).

A term closely related to the concept of attention is *consciousness*. In this work, this work will refer to the subjective state of mental content (such as perception, thoughts, and feelings). More precisely, being *conscious* means that something is represented in individual subjective experience, and is potentially available for use in further processes. "From a meta-cognitive perspective, mental content can stand in one of three relations to consciousness. It can be genuinely unconscious. It can be 'experientially conscious', that is, existing in the ongoing experience without being reflected on. It can be 'meta-conscious' and be explicitly represented as a content of one's own consciousness" (Winkielman & Schooler 2009: 52). Thus, consciousness differs subtly from attention, because not all content that exists in our ongoing experience need to be in the metaconscious focus of attention. Yet conscious mental content is *accessible*, that is, available to verbal report and higher-level processes, such as judgment, reasoning, and deliberate decision making. The distinction between consciousness and unconsciousness is often related to the preferential activation of the different processing routes. In essence, consciousness enables higher-order processing of information (logical inference, reasoning, calculus), whereas unconsciousness is associated with the fast and parsimonious "heuristic" mode (ibid.). This distinction seems to be justified insofar as operations of the elaborate route always involve operations on working memory, to which individuals have conscious access. Nevertheless, the simplification "conscious=rational, unconscious=automatic" is misleading, because automatic processes are also involved in supplying information to the working memory when the elaborate route is taken (Evans 2009).

Lastly, the term *intuition* generally describes the idea that judgments and decision making occur with little consciousness of the underlying processes (Strack & Deutsch 2009). The concept of intuition is commonly connected to a dominance of affect and "gut-feelings" in subjective experience (Kahneman & Frederick 2002), but also to simplifying processes such as categorization, stereotyping, automatic pattern recognition, and the use of rules of thumb (Glöckner & Witteman 2010). As Strack and Deutsch conclude, "although intuition is an idea that everybody appears to understand (at least intuitively), the meaning of intuition as a scientific concept is less clear ... intuitive judgments can be described by both the simplifying processes and their accompanying subjective experiences or feelings" (2009: 179f.). While psychologists direct their attention to underlying processes, the everyday meaning of intuition points towards the accompanying phenomenal experience. All in all, there is agreement that a multiplicity of different autonomous processes are responsible for creating what we commonly experience as intuition, subsuming many implicit operations of the cognitive system, to which we have no conscious access. Only the outcome of such implicit processing "pops up" in our consciousness and may (or may not) enter the focus of attention. Thus, the term *intuitive* indicates that judgments and decisions directly reflect impressions generated by a selective activity of automatic cognitive processes outside of consciousness, rather than being based on a more systematic reasoning process (Kahneman 2003).

Considering the proposed two processing modes and their underlying cognitive systems, researchers have come up with a host of terminological labels (of "near epidemic proportions," Evans 2008) to emphasize their difference. The following table summarizes the semantic diversity that authors have introduced in describing the distinct systems (table 1):

Table 1: Terminology of the dual-process paradigm

Processing Modes		Author
Heuristic	Systematic	Chaiken1980
Intuitive	Extensional	Tversky & Kahnemann 1983
Peripheral	Central	Petty & Cacioppo 1986
Automatic	Controlled	Bargh 1989
Categorization	Individuation	Fiske & Neuberg 1990
Theory-driven	Data-driven	Fazio 1990
Experiential	Rational	Epstein 1996
Automatic	Rational	Esser 1996
Associative	Rule Based	Sloman 1996
System 1	System 2	Smith & DeCoster 2000 Stanovich & West 2000
Impulsive	Reflective	Strack & Deutsch 2004

Despite the terminological variety, there is considerable agreement on the characteristics that distinguish the two processing modes (Kahnemann 2003). Authors agree that processing via the heuristic, intuitive, and associative *automatic mode* and its underlying cognitive system takes almost no effort, and is usually marked by the absence of consciousness; that is, its op-

eration and activity do not come into the focus of attention; they are neither intentional nor under the deliberate control of the actor. Interpretation, the subjective definition of the situation, and choice rely on simple heuristics and rules of thumb, guided by situational cues and the routine application of learned schemata, scripts, and routines. As Smith and DeCoster (2000) argue, the automatic mode draws directly from the human slow-learning memory system, which stores information gradually and incrementally, so that general expectancies, based on average, typical properties of the environment, can emerge. That is, the knowledge stored in the slow-learning memory is highly associative, "schematic," and concerned with regularities. As such, automatic processing "operates essentially as a pattern-completion mechanism" (ibid. 110), based on the retrieval, association, and categorization of similarities between stored knowledge and salient cues from the environment. Pattern completion via the slow, associative memory-system can also recall affective responses and evaluations associated with an object, and thus automatically activate attitudes.

In short, the automatic mode is "fast, effortless, associative, implicit, slow-learning and emotional" (Kahnemann 2003: 698). Answers provided by the automatic route and the associative cognitive system simply "pop" into the head, and may not seem to have any justification other than intuition—they become part of the stimulus information, rather than being seen as part of the perceiver's own evaluation or interpretation (Smith & DeCoster 2000). Therefore, the automatic activation of schematic knowledge structures has a great potential to affect judgment and decision making. Empirically, accordant effects have been related to, for example, the automatic activation and application of attitudes (Fazio 1990a), stereotypes (Fiske & Neuberg 1990), probability judgments (Kahneman & Frederick 2002), the execution of routines and rules of thumb as standard solutions to reasoning problems (Sloman 1996), thinking and reasoning (Epstein 1991), and the evaluation of persuasive messages (Chaiken 1980).

More recently, researchers have proposed that the automatic mode comprises in reality a set of interrelated autonomous and highly specialized subsystems (Tooby & Cosmides 1992, Stanovich 2004, Evans 2008). For example, the social heuristics introduced above (FOF framing and SEH) can be considered part of the autonomous operation of the fast, automatic cognitive system. Generally, it includes "domain-general processes of unconscious learning and conditioning, automatic processes of action regulation via emotions; and rules, discriminators, and decision-making principles practiced to automaticity" (Glöckner & Witteman 2010: 6). The automatic cognitive system operates on neuronal structures which are evolutionarily older than those associated with its rational counterpart, and it is often regarded as providing a direct perception-behavior link (Bargh & Chartrand 1999, Dijksterhuis & Bargh 2001).

On the other hand, the systematic, extensional, controlled, *rational mode* and its underlying (and evolutionarily younger) cognitive system is marked by consciousness and selective attention to mental content and situational stimuli; it is not based on the intuitive use of heuristics

and routines, but rests on an explicit, elaborate, and controlled reasoning process, which is "constrained by the amount of attentional resources available at the moment" (Bargh 1989: 4). Furthermore, it involves the search, retrieval, and use of task-relevant information, accompanied by the application of abstract analytical rules. Processing in the rational mode can be described as a conscious, controlled application of "domain-general" rules, abstract thinking and reasoning. In short "it allows us to sustain the powerful context-free mechanisms of logical thought, inference, abstraction, planning, decision making, and cognitive control" (Stanovich 2004: 47). However, it would be wrong to simply equate the rational route with logic: the concept of "systematic" processing is much broader, as it also includes, for example, the ability to engage in hypothetical thought and mental simulations, and it delivers an inhibitory function in that it monitors and suppresses influences from the automatic route (Evans 2008).

As Smith and DeCoster (2000) point out, processing in the rational mode uses both the slow and the fast memory system. In contrast to the slow memory system, the fast memory system is responsible for rapidly constructing new representations "on the spot," binding together information about novel aspects of experience in the particular context. It therefore allows reaction to new information and new situations. The short-term memory works as a "global workspace" into which both the automatic and the rational cognitive systems can "broadcast" their output, to enable a conscious reasoning process (ibid.), but this occurs at the cost of a relatively slow and serial operation. Moreover, humans can intentionally access and process previously stored knowledge to modify and refine judgments and decisions, for example when carefully evaluating a new persuasive argument, or when using individuating information to form a specific expectation of trustworthiness during an interaction. Taken together, the rational route is "slow, serial, effortful, more likely to be consciously monitored and deliberately controlled" (Kahnemann 2003: 698). A primary function of the underlying cognitive system is "to monitor the quality of mental operations and overt behavior" (ibid. 699), which implies that "doubt is a phenomenon of System 2" (ibid. 702), i.e. it is intrinsically tied to activation of the rational mode.

Apart from postulating those two different modes of information-processing, dual-process theories are also concerned with the interplay of the automatic and rational systems. Although there is still considerable debate (see Smith & DeCoster 2000, Evans 2008, Evans & Frankish 2009), some important insights have emerged. Importantly, "highly accessible impressions produced by System 1 control judgments and preferences, *unless modified or overridden* by the deliberate operations of System 2" (Kahnemann 2003: 716, emphasis added; see also Haidt 2001, Stanovich 2004, Strack & Deutsch 2004 and Thompson 2009). In other words, the automatic mode and its cognitive architecture are active "by default," while the systematic, rational cognitive system is activated in addition to this only when it is required to intervene in, correct, or support the operations of the intuitive system.

As pointed out before, the capturing of attention and the activation of the rational cognitive system can happen in either a passive or an active fashion, and this is highly context-dependent. Evolutionary adaptive heuristics, such as friend-or-foe detection, automatic facial recognition, generic monitoring of the audio-visual field for unexpected stimuli, the detection of pattern mismatches between stimuli and stored knowledge, and the use of internal signals such as (negative) emotions can all trigger the intervention of the rational system. Theoretical support for such a "default-interventionist" (Evans 2008) conception of dual processing comes from an evolutionary perspective: assuming that an elaborate reasoning process is time consuming and energy intensive, it is highly adaptive for an organism to fall back on "fast and frugal" procedures whenever they deliver an appropriate solution to a problem, allowing at the same time scarce cognitive resources to be saved (Gigerenzer & Goldstein 1996, Gigerenzer & Selten 2001, Gigerenzer & Gaissmaier 2011).

A number of dual-process models explicitly relate information-processing modes to overt behavior and action (Bargh & Barndollar 1996, Bargh et al. 1996, Dijksterhuis & Bargh 2001, Strack & Deutsch 2004), assuming that both direct and indirect links between perception and behavior exist. For example, Strack and Deutsch point out that "in the reflective [rational] system, behavior is guided by the assessment of a future state in terms of its value and the probability of attaining it through this behavior. In the impulsive [automatic] system, a behavior is elicited through the spread of activation to behavioral schemata" (2004: 229).[2] Thus, in the automatic mode, perception is directly connected to behavior, building on the spreading activation of stored knowledge structures and the resulting activation levels. At the same time, the engagement of the rational mode exercises an inhibitory function on impulsive responses, and allows for a choice which approximates a rational decision-making process. According to Fazio, "the critical distinction between the two models centers on the extent to which the behavioral decision involves effortful reasoning as opposed to spontaneous flowing from individuals' definition of the event that is occurring" (Fazio 1990a: 91).

Therefore, when the associative system delivers a sufficiently useful solution to a task (i.e. of interpretation or choice), and the fit between stored knowledge and perceptual input is high, it is not necessary to override or intervene with more effortful cognitive processes. Taken together, dual-process models demonstrate the notion of human bounded rationality as adaptive, highly flexible, and responsive to situational constraints. Note the striking similarity between the default-interventionist interpretation of dual-processing and sociologists' theories of everyday routine based on taken-for-grantedness (i.e. Schütz & Luckmann 1973). As pointed out

[2] They define a "behavioral schema" as an associative cluster that "binds together frequently co-occurring motor representations with their conditions and their consequences ... behavioral schemata and their links to other contents in the impulsive system can be understood as habits" (Strack & Deutch 2004: 229). Importantly, behavioral schemata can be easily activated by automatic processes, and perceptual input can automatically activate elements in the associative memory system linked to behavioral schemata.

previously, situations appear problematic and interrupt the routine only to the extent that the available knowledge is not sufficient to define them, that is, when "coincidence between the actual theme and the potentially relevant elements of knowledge does not occur sufficiently for the mastery of the situation in question" (Schütz & Luckmann 1973: 236)—in the terminology of the dual-process framework, this amounts to saying that in problematic situations, the automatic pattern-matching process of the associative cognitive system fails. However "with routine coincidence, 'interpretation' is automatic. No explicitly judging explication occurs in which, on the one hand, the situation and, on the other hand, the relevant elements of knowledge come separately into the grasp of the consciousness to be compared to one another" (ibid. 198).

It is natural to transfer the framework of dual-processing to the problem of interpersonal trust. Suspension and the "leap of faith," the question of conditional versus unconditional trusting strategies, the emergence of institutional, rule-based forms of trust, contrasted with the idea of a trust as a rational choice—from a dual-processing perspective, many of these aspects can essentially be answered in terms of the preferential activation of distinct cognitive systems, their information-processing state, and the degree of rationality involved in interpretation and the choice of a trusting act. The decisive question is whether the default mode of associative pattern recognition, paired with a routine application of trust-related knowledge, occurs smoothly and without interruption, or whether internal or external events trigger an activation of the elaborate rational mode. The dual-process paradigm provides a promising avenue for a broad conceptualization of interpersonal trust: it emphasizes that human rationality is bounded and variable, and more concretely specifies how such bounded rationality is to be understood. Importantly, it suggests that actors can use either an automatic or a rational mode of information processing in a trust problem, and provides a tool which allows trust researchers to incorporate the broad phenomenological foundations of trust. Unconscious, associative routines in judgment and decision making can prevail in situations that do not call for activation or intervention of the rational system. In such cases, actors may not be conscious at all of the trust problem, which explains the notion of "blind" trust, paired with an absence of doubt and of consciousness of the vulnerabilities involved. On the other hand, a trust problem may also be approached in terms of a thoughtful reasoning process and approximate a maximizing decision, linked to an activation of the rational mode and the underlying cognitive system.

4.2.2. Context Dependence

The idea of a contingent, flexible use of different processing strategies in a trust problem points to the significance of the context and the surrounding social environment for our understanding of interpersonal trust. As we have seen, the context of a trust relation defines the relevance of trust-related knowledge; it is central to interpretation, expectation formation, and the choice of a trusting act. A dual-processing perspective suggests that the context of the trust

relation also influences the preferential activation of the automatic or rational mode of information processing and the underlying cognitive systems. Both effects must be considered simultaneously when thinking about interpersonal trust.

The context-dependence of perception, judgment, and choice has been known in the social sciences for a long time, and it is commonly referred to as "framing" (Tversky & Kahneman 1981, 1986, for reviews see Kuhberger 1998, and Levin et al. 1998). The term was initially used in a strict sense to denote inconsistencies in human decision making—"framing effects" circumscribe the empirical observation that the wording of experimental instructions, the formulation of a decision problem (that is, its "framing") and other apparently superficial changes in presentation all exert systematic effects on judgment and choice. According to Tversky and Kahneman, the reason for these effects is a change in the way in which the decision-maker interprets the situation.[3] Framing effects were initially regarded as a shortcoming of the human mind; the cognitive heuristics proposed in the attempt to explain them were regarded as a "bias" to rationality.[4] But as our understanding of human cognition has advanced, researchers have reinterpreted the meaning of "framing effects," no longer regarding them as flaws, but instead conceiving of them as adaptive and "ecologically rational" (Allison & Messick 1990, Gigerenzer 2000). In a broad sense, a framing effect refers "to an internal event that can be induced not only by semantic manipulations but may result also from other contextual features of a situation and from individual factors" (Kuhberger 1998: 24). It is, in short, a synonym for the context-dependence of human cognition and the definition of the situation. Framing effects demonstrate that human actors flexibly use the signals available in the environment during the process of defining the situation and their subsequent choice of action.

By influencing the subjective definition of the situation, the context of a decision problem influences perception of others, as it "sets the frame" in which a person's behavior is evaluated and judged (Kay et al. 2008). For example, when the context indicates the relevance of a social norm, actors commonly define the situation accordingly, so that, on the level of overt behavior, norm compliance becomes more likely and observed behavior of others will be evaluated relative to the norm (Bicchieri 2002, 2006). Framing effects have been consistently (re)produced in many judgmental and choice tasks, including prisoner's dilemma situations (Deutsch 1973, Liberman et al. 2004), public good games (Sonnemans et al. 1998, Cookson

[3] "We use the term 'decision frame' to refer to the decision-maker's *conception of acts, outcomes and contingencies* associated with a particular choice. The frame that a decision maker adopts is controlled partly by the formulation of the problem and partly by norms, habits, and personal characteristics of the decision maker" (Tversky & Kahnemann 1981: 453, emphasis added).

[4] In their article, Kahnemann and Tversky (1981) introduce three specific cognitive heuristics in an attempt to explain the susceptibility of humans to framing effects in judgment and decision making under risk: the *availability heuristic* (people assess the probability of an event by the degree to which instances of it are "readily available" in memory), the *representativeness heuristic* (the likelihood of an event is assessed by its similarity to stereotypes of similar occurrences) and the *anchoring heuristic* (judgment is based on some initial value, or "anchor," from previous experience or social comparison and adjustments from that value from experience).

2000), and trust games (Burnham et al. 2000). All in all, researchers have convincingly demonstrated that human decision making is highly context-sensitive, and that humans readily respond to subtle contextual cues such as "the name of the game" (Kay & Ross 2003, Liberman et al. 2004), the presence or absence of material objects associated with particular social environments (Kay et al. 2004), the formulation of decision problems in terms of gain and loss frames (Tversky & Kahneman 1981, 1986, Andreoni 1995, Keren 2007), and many more.

A second psychological research paradigm that has extensively studied the prevalence of the *automatic* effects of situational cues to judgment, choice, and performance is the so called "priming" paradigm (see Bargh & Chartrand 1999, Wheeler & Petty 2001). Generally, priming research has been concerned with the unconscious activation and automatic use of stored knowledge structures, such as stereotypes, heuristics, scripts, schemata, and social norms, by presenting ("priming") them in unrelated tasks, often even subliminally. As Higgins (1996) points out, priming essentially operates as a manipulation of construct accessibility: the situational stimuli presented automatically trigger a spreading activation of the stored cognitive constructs. The constructs, once primed, are readily used by humans in consecutive tasks, and influence judgment and decision making in construct-consistent ways. In sum, "it is now accepted as common knowledge that exposure to specific trait constructs, actual behaviors, or social group members (whose stereotypes contain trait and behavioral constructs) can result in the nonconscious expression of the activated behaviors" (Wheeler & Petty 2001: 212). For example, if a stereotype about elderly people is activated, then subjects walk more slowly; likewise, exposing subjects to words related to rudeness versus politeness has assimilative consequences on behavior in subsequent discussions (Bargh et al. 1996). Importantly, environmental cues can also automatically activate motivational states and behavioral goals (Bargh & Chartrand 1999, Förster et al. 2007, Förster & Denzler 2009). When subjects are made aware of the primes, their influence disappears, which demonstrates the controlled intervention of the rational system in otherwise automatic processes (Higgins 1996). Priming research has demonstrated the automaticity of judgments and behavior in domains such as social perception (Baldwin et al. 1990, Andersen et al. 1996), stereotyping (Devine 1989), emotional appraisal (Lazarus 1991a), persuasion (Chaiken et al. 1989), and attitudes and judgment (Greenwald & Banaji 1995). The results of framing and priming research jointly point to the central role of the context in the activation and use of stored knowledge. Priming theory suggests that the environment and situational cues may set in motion automatic processes that influence the definition of the situation and behavior, without any conscious awareness on the part of the decision-maker.

To comprehend the importance of these findings for our understanding of interpersonal trust, let us reinspect the different theoretical approaches to trust discussed in the last chapter for their standpoint on context-dependence. For example, the proposition of basic trust and its

merging into a generalized and stable disposition to trust as a personality trait were qualified by the finding that their influence varies with "situational strength"—their influence will be high only when strong cues indicating trustworthiness are absent (Gill et al. 2005). Developmental models of trust naturally imply that trustor and trustee can identify and make use of cues that indicate the relevant trust-related knowledge, for example to guide the contingent use and activation of relational schemata. As pointed out by Gambetta (1988a), the threshold for favorable expectations of trustworthiness varies both in accordance with subjective and objective (contextual) circumstances, such as stake size. With respect to perceived characteristics of the trustee, Mayer et al. point out that "the trustor's perception and *interpretation of the context* of the relationship will affect both the need for trust and the evaluation of trustworthiness" (1995: 727, emphasis added), and they relate contextual factors to attributed characteristics of benevolence, ability, and integrity. Context-dependence is implicit in most sociological approaches emphasizing the importance of social embeddedness and the impact of social norms and culture on the build-up of trust. In these accounts, the effectiveness of institutions and structural assurance not only rests on system trust, but crucially depends on their situational salience and appropriateness. Likewise, context is most relevant in accounts focusing on situational normality, taken-for-grantedness and corresponding routine in the choice of a trusting act.

Rational choice models of trust (take, for example, Coleman's model) assume that generalized expectations of trustworthiness p are replaced by specific expectations p^+ in cases where individuating information is accessible. Without further elaboration, these accounts maintain that expectation formation is context-dependent, in that varying cognitive knowledge structures are activated and become situationally relevant. Moreover, extensions to the standard trust game, such as psychological games and models of social preferences, are developed on the basis of an exogenously given set of preferences and "initial beliefs." Yet as Dufwenberg et al. (2011) empirically demonstrate, initial beliefs are highly context-dependent—one primary effect of the context can be found in a shift in first and second-order beliefs. In other words: a change in the context influences those variables which are exogenous to the economic models. The authors conclude that, "framing effects can be understood as a two-part process where (i) frames move beliefs, and (ii) beliefs shape motivation and choice" (ibid. 14).[5] Contextual framing effects have been interpreted in the economic framework as determining the *reference points* involved in evaluating other players' intentions and their fairness or equity concerns. But, as previously indicated, empirical evidence suggests that preferences and utility functions themselves may depend on context, and may change in response to the environment

[5] Thus, guilt-aversion and reciprocity models may be an adequate formal representation of a given set of initial beliefs, yet they cannot account for the more important aspect that looms over them: the origin of "initial beliefs." In economic models, the context-dependence of trust is accounted for by a change in initial beliefs, which have a decisive impact on the strategies played.

and the stimulus context (Mellers et al. 1998: 457, Fehr & Hoff 2011). All in all, across discipline borders, context-dependence is an ever-present (although sometimes only implicit) element of trust theorizing.

For the whole enterprise of trust research, Ostrom declares that "the most immediate research questions that need to be addressed using second-generation models of human behavior relate to the effects of structural variables" (2003: 63), notably the impact of the physical, cultural, and institutional environment conveyed to the trustor in the form of situational cues. Yet although the importance of the context for interpersonal trust has been regularly recognized by trust researchers, historically and "across intellectual traditions, scholars have given limited attention to the role of the [social] context" (Lewicki et al. 1998: 441). In short, while its importance is never denied, the elaboration of context-sensitive models has remained elusive.

Unsurprisingly, trust researchers have more recently started to emphasize the role of "situated cognition" in our understanding of interpersonal trust (e.g. Kramer 2006, Nooteboom 2007). Kramer introduces the "intuitive social auditor" model, according to which "it is assumed that individuals possess various kinds of cognitive and behavioral rules to use when (1) trying to make sense of a given trust dilemma situation and (2) decide how to react on the basis of the interpretation they form of the situation" (ibid. 71). In this process, the trustor uses "orienting rules" to help decode and categorize a trust problem prior to action, "interpretation rules" to interpret the response of the trustee, and "action rules" representing "beliefs about what sort of conduct is prudent and should be employed in a trust dilemma situation" (ibid.). These rules include and reflect the various cognitive knowledge structures that people use to navigate through trust problems: "People's mental models include their *social representations*, which encompass everything they know about other people, including all of their trust-related beliefs and expectations, their *self-representations* ... and their *situational taxonomies* (e.g. their beliefs about the various kinds of social situations they are likely to encounter in their social lives)" (ibid. 82, emphasis in original). Kramer argues in favor of bounded-rationality and heuristic-processing approach, arguing that the application of orienting and interpretative rules is automatic and relatively mindless, if features of the situation are familiar and the context seems routine. Unfortunately, he does not further develop these propositions into a tractable theoretical model. However, his assertions are highly reasonable from the dual-processing standpoint. As previously argued, trust is often envisioned as being "blind" and unreflective, characterized by an absence of doubt. Given that doubt is a feature of the rational system (see Kahneman 2003, Evans 2008), the idea of unconditional, "shallow" trust points to the use of an intuitive, automatic mode of information processing in trust problems and during the choice of a trusting act.

A related point has been made by Huang and Murnighan, who propose that the beginnings of trust development "may occur beneath our conscious radar, via automatic, non-conscious

cognitive processes" (2010: 63). They experimentally show that subliminally priming relational schemata of relatives and close friends influences trusting behavior towards strangers. Their research addresses identification-based unconditional trust, and it demonstrates that the priming and subsequent automatic application of trust-related constructs can transfer even to unfamiliar contexts. A similar argument can be made with respect to the activation of other trust-related schematic knowledge structures, such as social norms, roles, and routines. They are often applied in a relatively automatic fashion, and point towards an automatic mode of information processing in trust problems. This perspective would also fit to the distinction between "affective-based" and "cognition-based" types of trust. As we have seen, the automatic mode is often characterized as impulsive, intuitive, and emotional. For example, affective states associated with interpretive schemes are likely to become activated along with the particular mental model, thereby "rounding up" the trustor's subjective experience. At the same time, they reassure the trustor of a continued reliance on the automatic mode in the case of positively valenced affective states. On the other hand, the arousal of negatively valenced affective states inhibits a direct cognition-behavior link, and triggers rational-system intervention and doubt. Such a state of the cognitive system is presumably connected to the emergence of types of conditional trust or distrust.

All in all, from a dual-processing perspective, situational cues that are associated with stored trust-related knowledge can be assumed to be highly decisive in determining the mode of information processing in a trust problem, the type of activated trust-related knowledge, and, as a result, the type of trust we can expect. Context fulfills a double function in this conceptualization of trust. First, it determines the relevance of trust-related knowledge and influences the definition of the situation. In this respect, the impact of the social environment on the emergence of interpersonal trust is often emphasized (Deutsch 1973, Mayer et al. 1995, McAllister 1995, McKnight et al. 1998, Kramer 2006, Keren 2007). Trust researchers assume that trustors can readily extract the relevant situational features which allow for the appropriate definition of the situation, which is the first step in the trust process—as Möllering notes, "the state of expectation needs to be understood as the 'output' of the trust-process ... it may become function 'input' for actions (risk-taking, cooperating) and associations (relationships, social capital) which in themselves, however, should not be confounded with trust...the process of trust ends with a state of expectation and begins with interpretation" (2001: 415).

But secondly, and even more importantly, the context influences the degree of rationality involved in solving a trust problem. The human cognitive system directly builds on perceptual input in regulating the mode of information processing. Obviously, when taking into account human cognitive architecture, the process of trust may begin even before a conscious and deliberate interpretation of the situation has been made, and without the inclusion of effortful reasoning or decision-making processes. This is the case when the automatic mode of infor-

mation processing is furnished by salient and appropriate situational cues. If the routine of everyday behavior can be maintained by successful pattern recognition and by the "matching" of situational stimuli with preexisting stored interpretive schemes, then the allocation of attention, the conscious awareness of trust problems, and doubtful reasoning processes about the choice of a trusting act may be fully absent.

4.2.3. Heuristics and Mental Shortcuts

The preceding sections have highlighted various routes by which humans can take shortcuts to judgment and decision making in a trust problem. Essentially, these shortcuts demonstrate that variable levels of rationality can be involved in the choice of a trusting act. Conceptually, they are often linked to the activation of the automatic route to information processing; in effect, stored associative knowledge structures and heuristics become a basis for unconditional trust. Some of these heuristics may be "hard-wired" (FOF, SEH); others may be learned over time (for example, generalizations such as frames, scripts, or stereotypes). Their unifying characteristic is that they relieve individuals of the need to approach the trust problem in terms of an effortful, systematic, and maximizing decision, and they can be applied to solve a trust problem automatically and without much conscious effort.

However, when we look closer, we inevitably encounter a confusing variety of such possible shortcut routes to trust—in fact, there is not only one way of solving a trust problem automatically and heuristically. As stated at the outset of this chapter, theoretical accounts which specify only one heuristic mechanism are necessarily incomplete. The important lesson that can be taken from adopting a dual-processing perspective in trust research is that adaptive rationality itself must be regarded as a basic dimension of the trust concept. We cannot think trust without thinking adaptive rationality. And when doing so, we have to concede that the "automatic" part of decision making is as multifaceted as is its rational counterpart, where the decision problem faced by a maximizing decision maker may take on a variety of specifications (see chapter 3.3).

In fact, the term "heuristic" has been used in the literature in various not necessarily consistent ways, and to date, there has been an abundant number of proposed mechanisms and processes which are regularly subsumed under the label "heuristic" (see Gigerenzer & Gaissmaier 2011 for a comprehensive review). The picture is complicated by the fact that there are competing ideas of how heuristics should be defined, and how they relate to the processing modes. For example, Gigerenzer & Gaissmaier define a heuristic as "a *strategy* that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods" (2011: 454, emphasis added). This view references the paradigm of

adaptive decision making introduced by Payne et al. (1993), who collected and worked out a number of heuristic decision-making strategies.[6] Notably, this definition is quite narrow because the concept of a heuristic refers exclusively to simplifying *choice rules*, which are applied in a relatively controlled fashion in choice problems. We will have to add the important point that heuristics can also be applied with the goal of arriving at an interpretation and a subjective definition of the situation more quickly, frugally, or accurately than with more complex methods. For our purposes, then, the term "heuristic" cannot be limited to simplifying choice rules.

A much broader definition is proposed by Chaiken et al. (1989), who define heuristics as "learned knowledge structures that may be used either self-consciously or non-self-consciously by social perceivers" (ibid. 213). These knowledge structures include declarative and procedural knowledge, such as frames and scripts, all of which may be used to simplify a task such as interpretation, judgment, or decision making.[7] In this regard, choice rules are merely a special case. Generally speaking, many forms of trust-related knowledge, such as relational schemata, generalized expectations, roles and norms, schematic knowledge of situations (frames, or "situational taxonomies," in Kramer's words) and behavioral scripts can be used as heuristics to simplify a trust problem. Importantly, "when processing heuristically, people focus on that subset of available information that enables them to use simple inferential rules, schemata, or cognitive heuristics to formulate their judgments" (ibid. 213). In other words, heuristic processing is largely based on the heuristic *cues* available in the environment; interpretation, judgment, and choice are accomplished by using available knowledge, instead of by relying on a more detailed analysis of information.

One complicating factor is that heuristics can be used in both an automatic and a rational fashion (Chen & Chaiken 1999, Kahneman & Frederick 2002). On one hand, individuals need not necessarily be aware of their use of heuristics —only the heuristic cue that leads to the activation of the heuristic, and the result of its application are part of conscious experience. For example, merely seeing a doctor in professional clothing (a cue) may be sufficient to trigger the use of a judgmental heuristic, such as "doctors are competent and trustworthy" (rule), which influences judgments of trustworthiness (result). That is, "although heuristic processing entails, minimally, an awareness of a heuristic cue in the environment, this does not imply that

[6] For example, the *lexicographic* heuristic (select an alternative which is best in terms of the most important attribute, i.e. "take the best"), the *equal weights* heuristic (ignore probabilistic information), the *satisficing* heuristic (consider one alternative at a time in their natural order, and select an alternative if its attributes reach an "aspiration level"), or the *elimination by aspects* heuristic (determine the most important attribute, eliminate all alternatives that do not reach a threshold, and continue with the next attribute until one alternative is left).

[7] Broadly speaking, *declarative* knowledge is stored knowledge of facts and events. It is often symbolically coded, associative, and it can be consciously accessed. On the other hand, *procedural* knowledge is tacit knowledge of "how to do things". It includes the skills we have learned, and it cannot be expressed directly. Procedural knowledge also includes habits and routines of everyday behavior.

perceivers are necessarily aware of the activation of a corresponding heuristic that occurs as a result of encountering this information, or of their application of this rule to their current judgmental task" (Chen & Chaiken 1999: 86). On the other hand, heuristics can also be used in a controlled fashion; one prominent example is the choice rules referred to above, which can be applied in a rational or an automatic fashion. Although the idea of using heuristics is often linked to the automatic mode of information processing and the activation of the underlying fast, associative, and "intuitive" cognitive system (e.g. Fazio 1990, Strack & Deutsch 2004), heuristics can also be applied in a controlled reasoning process. Yet as Chen and Chaiken (1999) argue, the larger share of our day-to-day heuristic processing is in fact automatic and unconscious.

Following the above definition, heuristics are elements of learned knowledge. However, s*ubjective experiences* may simultaneously serve as heuristic cues, and also as judgmental heuristics. They present an exceptional case of mental shortcuts which do not fall directly into the scope of the definitions given above, although they have been regularly described as heuristics.[8] According to Schwarz and Clore (1996), the affective, cognitive, and bodily states of an individual form an important part of his subjective experience, and serve as signals that influence the way in which information is processed. We have already looked at the "affective" aspect of this proposition in chapter 2.2.4. Affective experiences, such as mood and emotions, can have a direct influence on the processing mode; negatively valenced affective signals which indicate a problem foster vigilance and the adoption of detail-oriented elaborate processing, whereas benign signals promote an automatic processing mode (Schwarz 1990, Bless & Fiedler 2006). But at the same time, affective feelings can serve as a heuristic in their own right ("affect heuristic", Slovic et al. 2002). Individuals use affective feelings as a source of summary information, qualitatively different from stored knowledge, in order to judge a target. In the context of interpersonal trust, this was termed the "feeling of rightness" involved in a trusting act, indicating that trustors rely on their currently perceived affective state as summary information to judge the trustworthiness of a trustee. By asking themselves "How do I feel about it?" emotions are often used as experiential heuristic information to form a variety of judgments (Forgas 2002). As Chen and Chaiken (1999) argue, subjective experiences are particularly prone to being used automatically, or "intuitively."

This also holds for "cognitive" experiences such as ease of retrieval, processing fluency, the feeling of knowing, and familiarity (see Bless et al. 2009, Greifeneder et al. 2011). Cognitive experiences indicate whether or not the cognitive apparatus is working smoothly—the internal

[8] As Strack and Deutsch point out, intuitive judgments "use cues that are less complex, which can be found *either* in the environment *or* as an internal response to the environment, such as affective and non-affective feelings, conceptual activation, and behavioral responses" (2009: 190). Thus, a dual-system perspective suggests that various and "potentially very different processes generate the simplifying cues that may feed into judgments" (ibid.), although all of them can contribute to the more general experience of "intuition."

functioning of the mind and the ease or difficulty with which processing occurs may also become a subjective experience. As with affective experiences, cognitive experiences can have a pronounced impact on individual information processing. Disruptive experiences which signal the presence of problems in processing (low fluency, difficult retrieval, unfamiliarity) are likely to trigger a systematic mode of information processing. Similar to affective feelings, cognitive feelings are used as heuristic summary information that influences the evaluation and judgment of targets. For example, targets are evaluated more positively whenever their stimuli can be processed fluently (Reber & Schwarz 1999, Reber et al. 2004), and individuals often make use of the ease with which information comes to mind as a substitute for content information in forming a judgment (Tversky & Kahneman 1973, Schwarz et al. 1991).

Cognitive experiences can be fruitfully connected to the theoretical concepts prevalent in the trust literature. For example, fluency and ease of retrieval experiences directly impact our sense of situational normality, and therefore relate to the build-up of trust. In this line, Greifeneder et al. (2010) empirically demonstrate that the experience of ease of retrieval influences the choice of a trusting act and the attributions of procedural fairness. When thinking about few (easy) or many (difficult) *unfair* aspects of a trust game, subjects tend to rely on ease of retrieval, in that a recall of few negative aspects (high fluency and ease of retrieval) results in *less* behavioral trust and in lower ratings of procedural fairness, whereas a recall of many negative aspects (low fluency and ease of retrieval) results in *more* trust and in higher ratings of fairness. Researchers have also accumulated evidence that processing fluency elicits positive affect. In other words, error-free processing "feels good" because it indicates a positive state of affairs within the cognitive system and the outer world (Winkielman et al. 2003). Thus, the heuristic use of cognitive feelings may be relevant to the build-up of interpersonal trust and, supposedly, it is especially important in initial trust formation and one-shot situations, where more specific sources of trust-related knowledge are unavailable.

The above sections suggest that a comprehensive definition of a *heuristic* would characterize it as "a learned knowledge structure or subjective experience which may be used either self-consciously or non-self-consciously by social perceivers to make interpretation, judgment, and choice more quickly than with more complex methods." Thus, when we use the term heuristic, we indicate that a task, such as interpretation, has been simplified internally by cognitive or experiential shortcuts. In this line, Kahneman and Frederick (2002) propose that attribute substitution—the reduction of complex tasks to simpler operations—is in fact the defining characteristic of heuristics: "judgment is said to be mediated by a heuristic when the individual assesses a target attribute of a judgment object by substituting another property of that object—the heuristic attribute—which comes more readily to mind" (2002: 53).

The relevance of heuristics to trust is clear: heuristics influence expectations of trustworthiness and alter the subjective experience of trust; if paired with an automatic processing mode,

they may even prevent a conscious elaboration of the trust problem. This explains the notion of unconditional trust as portrayed by psychological and philosophical trust researchers. Unconditional trust is regularly connected to the application of learned knowledge structures (generalized expectations, schemata, scripts), and to the preferential impact of subjective experiences (affect, familiarity). These two classes of heuristics, paired with the assumption of their automatic use, can quite generally account for those types of trust which are regularly denoted as unconditional (i.e. identification-based, affect-based, rule-based trust). Thus, "intuition" during the choice of a trusting act can have a broad phenomenological foundation, ranging from the swift application of relational schemata, rules, roles, or routines, to the "heuristic" use of affective and cognitive experiences and the preattentive influence of "hard-wired" fast and frugal heuristics, such as SEH or FOF. In the case of unconditional trust, it is characteristic that these heuristics are applied in the automatic mode. Otherwise, the results of the heuristic process are merely integrated into a controlled and systematic judgment; they may be called into question and revised. Such elaborate and controlled reasoning process is characteristic of conditional trust.

4.2.4. The Neuroscience of Trust

In a very recent development, trust researchers have used neuroscience techniques such as brain imaging, brain stimulation, the study of brain lesions, psychophysical measurements, and pharmacological interventions to study the neurobiological processes involved in trust (Zak 2007, Fehr 2009, Rilling & Sanfey 2011). The neuroscience of trust has emerged as one of the most important offspring of the more general and rapidly advancing field of "neuroeconomics" (Zak 2004, Camerer et al. 2005). Neuroeconomic research focuses on the physical substrate of the cognitive system—brain regions, neural circuits, neural activity, and so forth—to infer details about the black box of the brain and explore its functioning in individual behavior in social decision making situations. In short, neuroeconomics seeks to ground economic behavior in the details of the brain's functioning.[9]

The two broadest findings that this research field has contributed confirm the core tenets of the dual-processing paradigm: (1) human behavior is, to a large extent, automatic and (2) behavior is strongly influenced by finely tuned affective systems which intervene and interact with the deliberative system (Camerer et al. 2005). Interacting with humans and making decisions in a social context reliably activates areas associated with affect and emotions. What is more, there is ample evidence that social preferences have a "neural correlate." Cooperation, reciprocation, and the altruistic punishment of others activate neural circuitry that overlaps

[9] The following section presents a brief overview over the rapidly growing field of neuroeconomics without going into much detail. Excellent summaries and informative introductions to neuroscience, its terminology, and its methodology, can be found in Zak (2004), Camerer et al. (2005) and Rilling and Sanfey (2011).

closely with circuitry anticipating and representing other types of rewards (Fehr & Camerer 2007). At the same time, the activation of circuitry associated with negative emotional states, such as fear or disgust, can be observed in response to inequity, nonreciprocity, and the violation of expectations, both real and hypothetical (Sanfey 2007). Interacting with a real human, in contrast to with a computer, genuinely activates a number of areas associated with the "theory of mind" (Rilling et al. 2002). Furthermore, dealing with social uncertainty substantially differs from dealing with nonsocial risks—on a neural basis, social and nonsocial risks cannot be equated (Fehr 2009). Brain imaging studies have also substantiated the distinction between risk and ambiguity—which both activate different areas of the brain (McCabe et al. 2001)—and have revealed a number of specific neural circuits involved in the implementation of and the compliance to social norms, as well as in dealing with potential conflicts among norms (Spitzer et al. 2007). Even more intriguingly, a number of studies have explored the modulating effect of hormones such as testosterone and serotonin on neural structures, showing that they can dampen or excite brain activity, thereby strongly influencing behavior (Rilling & Sanfey 2011).

In these and many related studies, a number of regions of the brain which are regularly involved in social decision making and social interaction have been identified. Generally speaking, regions of the *prefrontal cortex* (PFC) are associated with control and inhibition of emotional impulses stemming from components of the automatic system, such as the *amygdala* (fear, betrayal aversion, processing of potential threats), the *anterior insula* (aversive responses to unreciprocated cooperation, norm violations, empathy), and the *striatum* (a mid-brain dopamine cell region which is speculated to provide the brain's general reward system). The *ventromedial prefrontal cortex* (VMPFC) is crucially involved in evaluating long-term benefits of cooperative relationships and abstract rewards, and in regulating emotional reactions. The *dorsolateral prefrontal cortex* (DLPFC) exerts cognitive control for overriding selfish impulses, and the *ventrolateral prefrontal cortex* (VLPFC) is involved in overriding aversive reactions to unfair treatment (ibid.). These regions have been found to be frequently involved in social interaction and, more importantly, during interpretation and choice in trust problems.

Neuroeconomic studies have unveiled a number of results that help to trace the emergence of trust back into the neural components of the automatic and rational system, and to the chemical and neural processes involved. To begin with, judgments of trustworthiness are directly related to automatic amygdala activation, with untrustworthy faces increasing activation levels, even when the judgment is made implicit (Winston et al. 2002). Consequentially, patients with amygdala lesions consistently overestimate other people's trustworthiness, suggesting that the role of the amygdala in processing potential threats and dangers extends to the domain of social interaction in trust problems (Adolphs et al. 1998). Another region that is crucially involved in the choice of a trusting act is the VMPFC. Lesions in this area result in less trust

in trust games (Krajbich et al. 2009). Since the VMPFC registers long-term benefits that could emerge from a successful trust relation, it potentially helps to surmount the immediate fear of betrayal associated with the decision to trust that stems from the amygdala (Sanfey & Rilling 2011). Krueger et al. (2007) have further identified the *paracingulate cortex* (PCC) and *septal area* (SA) regions as being involved in the choice of a trusting act. The PCC is a neural structure involved in mentalizing and inferring the mental states, feelings, and beliefs of others, while the SA is intricately connected to social attachment behavior. Using functional imaging to explore the neural activity in a trust problem, Krueger et al. (2007) also show that different trusting strategies—conditional and unconditional—result in the preferential activation of distinct neuronal systems. While unconditional trust selectively activates the SA, conditional trust selectively activates the *ventral tegmental area* (VTA), an area linked to the evaluation of expected and realized rewards. Conditional and unconditional trusting strategies differ not only with respect to their behavioral outcomes, but also with respect to their decision times, which become increasingly shorter over time for unconditional trust.

Furthermore, there seems to be a fundamental connection between neural activity in the SA and trust: the SA plays an important role in the release of the neuropeptide oxytocin (OT)—a key hormone involved in a number of complex social behaviors, such as maternal care, pair bonding, and social attachment. In a widely-received study, Kosfeld et al. (2005) exogenously manipulated OT levels and found that trust significantly increased in comparison to a control condition. At the same time, OT did not decrease risk-aversion in general, but its effects were limited to the social risks arising from interaction in the trust problem. In a follow-up study, Baumgartner et al. (2008) could replicate these results and, combining the design with neural imaging techniques, found that OT treatments reduce activity in the amygdala, mid-brain, and striatum areas, all of which are critical in signaling and modulating fear responses. They concluded that OT reduces fearful responses to the social uncertainty involved in trust problems, enhancing the subject's ability to overcome social uncertainty and choose a trusting act. Furthermore, OT treatments, although effectively manipulating trusting behavior, influenced neither measures of mood, calmness, or wakefulness (ibid.) nor the subject's expectations of trustworthiness (Kosfeld et al. 2005). In economic terms, this suggests that OT does not influence beliefs, but directly shapes social preferences, leaving more general risk and ambiguity aversion preferences unaffected (Fehr 2009).

Taking things together, the neuroscience approach to trust enables researchers to focus on the effect of particular brain structures on interpretation and choice in a trust problem. Essentially, it attempts to pin down those neural correlates of the automatic and rational systems which are crucially involved in the choice of a trusting act. The studies presented here suggest that there is a strong connection between trusting behavior and neural processes. As Zak and Kugler boldly put it, "trust is chemical" (Zak & Kugler 2011: 143). Even when one's developmental

history, prevailing social norms, and current events influence the trusting strategies a trustor adopts, they do so by modulating OT release, which, according to Zak and Kugler (2011), may potentially constitute the single causal pathway through which trust and trustworthiness can be explained. However, components of both the automatic and rational system are involved in influencing the judgment of trustworthiness and the choice of a trusting act. Therefore, on a more general theoretical level, it is important to know *when* the different neuronal structures are active and are determining trust. Having established and specified the principal routes by which trust can build up, it is now our task to be more precise about the determinants of the processing modes and the degree of rationality involved in the choice of a trusting act.

4.3. Determinants of Information Processing

The most immediate question that arises when thinking about interpersonal trust in terms of the dual-process notion is naturally, when exactly can we expect each mode to occur? When is the emergence of a certain "type" of trust likely? When are heuristics used automatically to solve a trust problem? In other words, how does the human cognitive system solve the problem of *mode selection* and adaptive rationality? Obviously, a number of factors determine the mode in which information is processed, and these pave the route along which the trust problem is approached. Both individual and situational factors have to be considered when thinking about the determinants of information processing. To date, researchers have offered a plethora of variables that potentially define, influence, and moderate the processing mode (see Chaiken & Trope 1999). These lists of "moderators" are often paradigm-specific, and generally portray the fact that cognition responds flexibly to the environment and to the task structure. However, given the fact that the human cognitive system is highly adaptive, it is likely that such lists can never be comprehensive or complete. As it is, many of the proposed variables can exert multiple effects under different circumstances, serving as cues, as information, as mere "biasing" factors, or at times as determinants of information processing; a variable that increases information processing in one context may decrease it in another, depending on intrinsic factors such as personal relevance or active goals (Petty & Wegener 1999). Principally, researchers argue that moderating variables have an impact on information processing via their influence on one of the four main determinants which have surfaced as central to the degree of rationality involved in interpretation and choice:[10] Shared by most dual-processing models is the proposition that (1) opportunity, and (2) motivation crucially determine whether information is processed in a more automatic or a more rational fashion (Smith & DeCoster 2000). In reviewing the existing literature, Mayerl (2009: 117) suggests a distinction between

[10] The following review is necessarily short and incomplete. For an extensive overview, the reader is referred to the volume edited by Chaiken and Trope (1999), who join articles by the most influential scholars of the field.

situational, *individual-intrinsic*, and *thematic* dimensions of opportunity and motivation. Moreover, most dual-process models emphasize (3) the accessibility of stored knowledge and its fit with situational stimuli as an important determinant (Higgins 1996, Kahneman 2003). Lastly, (4) the cognitive costs and efforts associated with different processing strategies are decisive in determining the degree of rationality and the decision strategies used (Payne et al. 1993).

4.3.1. Opportunity

The factor of *opportunity* emphasizes that the cognitive resources of humans are limited by both individual and situational constraints which may prevent the engagement of the rational mode. It refers to the available processing time and attentional resources, and denotes whether or not the opportunities to engage in rational information processing do actually exist. If opportunities do not exist, then the processing of information by the rational route is simply not feasible. According to Fazio, "situations that require one to make a behavioral response quickly can deny one the opportunity to undertake the sort of reflection and reasoning that may be desired" (1990a: 92). This aspect refers to the situational dimension of opportunity; it principally equates to available time and the presence or absence of time pressure.

Apart from that, opportunity involves an individual-intrinsic dimension, which refers to ability and cognitive capacity (Kruglanski & Thompson 1999). While *ability* denotes general cognitive skills and estimated "self-efficacy," that is, one's own judgment of how effectively information can be processed, *cognitive capacity* refers to the general availability of the scarce resource of attention and the temporary "cognitive load" (Shiv & Fedorikhin 2002) experienced by the decision-maker. As suggested by research on ego-depletion (Baumeister et al. 1998), even minor acts of self-control, such as making a simple choice, use up the limited self-regulatory resources available. Likewise, engaging with concurrent tasks at the same time drastically limits the amount of available cognitive resources. Whether or not an individual can access his or her cognitive resources, and whether or not they are "free" to use, is an important determinant of individual-intrinsic opportunity. Lastly, thematic opportunity pertains to the objective presence or absence of thematic knowledge with respect to a given decision problem, and the individual ability to make an appropriate judgment in a certain thematic domain (Eagly & Chaiken 1993).

4.3.2. Motivation

The *motivation* to engage in a controlled, systematic, rational mode of information-processing also has situational, individual-intrinsic, and thematic subfactors. Principally, the lack of motivation to engage in effortful reasoning inhibits the engagement of the rational mode. Situational motivation describes the perceived "importance" of a task or decision-problem, and the perceived "responsibility" for an outcome (Eagly & Chaiken 1993), as well as the "fear of in-

validity" (Sanbonmatsu & Fazio 1990) when making a judgment. It can be influenced, for example, by making judgments public or by having third parties or experts observe and evaluate the decisions. Furthermore, the prospective gains and losses involved in the situation, that is, the objective structure and the stakes of the decision are important determinants of situational motivation (Payne et al. 1993). Likewise, surprise and salient cues can increase the situational motivation to engage in more elaborated processing by passively capturing attention.

Psychologists have emphasized individual-intrinsic factors of motivation, such as "need for cognition" (Cacioppo & Petty 1982), "accuracy-motivation" (Chaiken 1980, Petty & Cacioppo 1986), and "faith in intuition" (Epstein et al. 1996). These are regarded as relatively stable personality traits influencing the tendency to engage in rational or automatic modes of decision making. In short, "people with a preference for intuition base most of their decisions on affect, resulting in fast, spontaneous decisions, whereas people with a preference for deliberation tend to make slower, elaborated, and cognition-based decisions" (Schunk & Betsch 2006: 388). Concerning thematic motivation, it is influenced by factors such as individual involvement, the personal relevance of a judgment or decision task in the specific thematic domain, and target ambivalence and security in judgment (Eagly & Chaiken 1993). Strack and Deutsch (2004) emphasize the role of affect as a situational and thematic-motivational determinant. Negative emotional reactions to stimuli often trigger the activation of the rational mode, while positive affective states promote a more "top-down" automatic processing (see also Bless et al. 1996, Bless & Fiedler 2006).

Moreover, goals and expectations influence the mode of decision making by influencing motivation. Active goals and expectations can foster more rational processing of information, for example, when a desired outcome calls for systematic elaboration, or when expectations direct attention towards a systematic analysis (Fiske 1993, Bargh et al. 2001, Molden & Higgins 2005). Likewise, they can also prevent a more detailed analysis and foster the automatic mode, if the current goals do not ask for an accurate decision. Goals harbor an individual-intrinsic and a situational-thematic dimension, as they are often context-specific, but at the same time they may be influenced by individual (long-term) values and higher-order goals which actors seek to achieve.

4.3.3. Accessibility, Applicability, and Fit

Accessibility is the ease (or effort) with which particular mental contents come to mind. As it is, "the accessibility of a thought is determined jointly by the characteristics of the cognitive mechanisms that produce it and by the characteristics of the stimuli and events that evoke it" (Kahneman 2003: 699). In a nutshell, "accessibility can be defined as the *activation potential* of available knowledge" (Higgins 1996: 134, emphasis added). Thus, accessible knowledge is capable of being activated and used, but it exists in a rather latent state. It is important to dif-

ferentiate accessibility from *availability*—that is, whether or not some particular knowledge is actually stored in the memory system. Availability is a necessary condition for accessibility: if availability is zero, then accessibility is zero as well (Higgins 1996). The general position is that "the greater the accessibility of stored categorical knowledge, the more likely that it would be used to categorize stimulus information" (ibid. 133). Thus, interpretation and the subjective definition of the situation, the way individuals make sense of a situation and form judgments, is based on that information which is most accessible at the moment. Situational stimuli which foster the activation of stored knowledge increase its *temporary accessibility*. If mental contents are temporarily accessible, then they readily come to mind and are activated and used during the processing of information—the results of framing and priming research introduced earlier exemplify this perspective.

But accessibility has an individual-intrinsic dimension as well, often referred to as *chronic accessibility*. While temporary accessibility is the source of context effects in judgment and decision making, chronically accessible information lends judgments and decisions some context-independent stability (Schwarz 2009). Importantly, researchers have shown that chronic accessibility increases the likelihood that knowledge is activated and used in a task or judgment. With respect to attitude activation, Fazio notes that "the likelihood of activation of the attitude upon mere observation depends on the chronic accessibility of the attitude" (1990: 81). That is, chronic information has a higher activation potential than nonchronic information (Higgins 1996: 140f.).

Remember that the automatic mode and its underlying cognitive system are often characterized as associative pattern recognition mechanisms. Whether pattern recognition succeeds or fails is an important trigger in rational system interventions—a situation appears problematic and calls for a systematic analysis to the extent that stored knowledge is *not* sufficient to master it. Thus, when thinking about the way in which the mode of information processing is determined, we have to consider not only the accessibility of stored knowledge, but also its applicability (or "fit") with respect to the perceptual input as a determinant of actual knowledge activation (Higgins 1996: 154f.).

To describe the degree of perceived overlap between stimulus data and stored knowledge and its applicability, we will henceforth use the term *match*. The more features of the stimulus are in line with the stored knowledge (the higher the match), the higher the likelihood that the construct will be activated and used to categorize the stimulus. A high match between stimulus and accessible stored knowledge can be sufficient to trigger an automatic behavioral reaction. For example, relating to the domain of attitudes, Fazio proposes that "behavior simply follows from a definition of the event that has been biased by the automatically activated attitude. Neither the activation of the attitude from memory nor the selective perception compo-

nent require conscious effort, intent, or control on the part of the individual" (Fazio 1990a: 84, see also Fiske & Neuberg 1990).

Some authors suggest that a high match is connected with cognitive fluency experiences, which go along with an automatic use of heuristics, even when the heuristic itself does not get a "grip on the mind." According to Thompson, "heuristic outputs are delivered into conscious awareness accompanied by a metacognitive experience that is largely ... determined by the fluency with which the output was retrieved" (2009: 177). In other words, cognitive experiences such as fluency and ease of retrieval can be interpreted as the experiential side of information processing, based on the accessibility, applicability, and fit of mental content. As Thompson furthermore argues, the strength of the cognitive experience of fluency is a key trigger of rational system interventions: with high fluency, interventions are unlikely to occur; with low fluency, the probability for a rejection of heuristic judgments and a controlled re-evaluation of information is high. In either case, the accessibility of stored knowledge structures and their match to situational features are the most important promoters of the automatic processing mode and the cognitive experiences that go along with it.

4.3.4. Effort-Accuracy Tradeoffs

Dual-process models assume that processing modes and decision strategies differ in the mental effort, or "costs," attached to them, and that a tradeoff between the anticipated *effort* incurred and the anticipated *accuracy* provided is made when selecting a decision strategy. For example, Payne et al. (1988, 1992, 1993) analyze different choice rules with respect to the necessary elementary information processes (that is, basic operational steps such as retrieving some value from memory, storing a value in memory, executing an addition, comparing alternatives on an attribute etc.). They show that the "weighted additive rule," a strategy resembling expected utility maximization, is by far the most cognitively effortful decision strategy available. On the other hand, simpler heuristic strategies are less costly in terms of cognitive effort, but they are also less accurate, with a random choice being the least effortful and least accurate method. The authors propose that effort-accuracy tradeoffs are the principal mechanism governing the contingent selection of decision-making strategies. Most dual-process theories agree in proposing such a "sufficiency principle." In short, individuals "will employ a systematic strategy when reliability concerns outweigh economic concerns, and a heuristic strategy when economic concerns predominate" (Chaiken 1980: 754).

What are the efforts, or "costs" that actors incur with processing? We can define *mental effort* as the number of attention-demanding operations needed to be executed in working memory in order to perform a task (Kahneman 1973). Mental effort is directly connected to physiological processes of energy mobilization, and can therefore be measured, for example, in terms of cardiovascular responses and neural activity (Fairclough & Mulder 2011). The difficulty of a

task is one important determinant of mental effort, as it dictates whether it is necessary to make attention-demanding computations (Mulder 1986). Empirically, researchers have shown that increases in task complexity can shift information processing to more heuristic strategies (Payne 1976, Heiner 1985). This suggests that effort-accuracy tradeoffs are involved in determining processing modes. However, such effects cannot be regarded in isolation from other determinants, in particular from intrinsic and extrinsic cognitive motivation. Increases in task complexity can also encourage more elaborate processing; actors may differ in intrinsic accuracy motivation and "need for cognition," which influences effort-accuracy tradeoffs. More generally speaking, social psychology has portrayed humans as "cognitive misers" (Fiske & Taylor 1991) and, in a refined metaphor, as "motivated tacticians" (Fiske 2004) who quickly use prior knowledge and cognitive shortcuts to avoid the effortful route of rational processing whenever it is affordable, but who nevertheless can flexibly alter the amount of processing involved in a judgment or decision whenever it is necessary to do so. The idea that effort is one important determinant of information processing is a recurring theme in most dual-processing accounts.

Concerning the interaction of the determinants, dual-processing models converge on some important points, although the precise way they interplay is still widely debated. First and foremost, there is widespread consensus that both opportunity and motivation are necessary conditions for selecting the rational mode (Chaiken 1980, Petty & Cacioppo 1986, Fazio 1990a, Strack & Deutsch 2004). An absence of either factor prevents the intervention of the rational system, simply because it is not feasible, or because it is not required. Second, as suggested by the accuracy-effort frameworks, the costs of using a more elaborate strategy are negatively correlated with accuracy. This tradeoff is directly reflected in an interactional pattern involving effort and motivation: according to the "sufficiency principle" (Chen & Chaiken 1999), perceivers attempt to strike a balance between minimizing cognitive effort and satisfying their current motivational concerns. While a high "accuracy motivation" to elaborate principally increases the likelihood of using the rational mode, high costs and effort may counterbalance and demotivate its usage.

Furthermore, accessibility and applicability are deemed to be of importance in most models: individuals, when processing automatically, rely on that content which is most accessible when performing a specific task. However, accessibility alone is not sufficient because the automatic activation of stored knowledge also depends on the match between the schema and stimulus input (Fazio 1990a, Fiske & Neuberg 1990, Higgins 1996). Thus, only when accessibility and applicability are highly matched, can the routine of pattern recognition be maintained and stored knowledge be applied automatically. On the other hand, a mismatch between perceptual input and accessible stored knowledge increases the motivation to intervene with the rational system, given that sufficient opportunities and motivation exist (Fiske et al.

1999). The default interventionist perspective here puts forward puts a special emphasis on the role of the match in determining the processing mode: as long as no problems occur and the situations encountered match the stored knowledge, the default of automatic routine in everyday behavior can be maintained, given that an actor is not motivated to engage in a rational elaboration.

4.4. Dual-Processing: A Critical Assessment

Taking things together, the dual-processing paradigm constitutes a valuable resource that can inform trust research because it demonstrates adaptive rationality as a fundamental characteristic of human (inter)action. This fact must not be taken lightly: presumably, adaptive rationality is involved in every choice of a trusting act and plays a role in every solution to a trust problem. Prior to the rise of the dual-processing paradigm, trust researchers pointed to trust as a "mix of feeling and thinking"; the results achieved in this area can now help us to better understand the determinants and influence of adaptive rationality, to improve the behavioral and phenomenological foundations of interpersonal trust (i.e. its subjective experience), as well as to lead the way to a causal explanation of different "types" of trust.

A most valuable conclusion that can be drawn in the face of dual-processing research is that *adaptive rationality must be regarded as a fundamental dimension of the trust concept* itself. Any attempt to explain trust with the use of only one "route" and without reference to the processing state of the cognitive system must necessarily remain incomplete. The neglect of adaptive rationality is one reason for the diverse and often conflicting ideal-type classifications discussed in the preceding chapters (i.e. calculus-based versus affect-based trust). In fact, these types can easily be integrated along the dimension of adaptive rationality. But trust theory shares with the dual-processing paradigm the dilemma that, although detailed specifications exist for each "type," the causal links between them—that is, the more general theory that would connect them—are missing.

It is worthwhile to note that dual-processing accounts are relatively silent when it comes to a precise explication of the interplay between the four fundamental determinants and their link to action. The research reviewed above has convincingly demonstrated the existence of individual adaptive rationality. It has gathered important insights about how the determinants influence the degree of rationality involved in interpretation and choice. But to date there has been no theoretical account available that unites all variables and explicates their interplay at the same time. Furthermore, the models proposed—in stark contrast to the paradigm of rational choice—do not offer explicit *selection rules* that would govern the definition of the situation and the selection of scripts and of actions, meaning that the actual link between cognition and action cannot be formally established (Esser 2000a: 239f., Mayerl 2009: 52f., 151f.,

2010). In other words, what is lacking is a tractable formal model of adaptive rationality, and its theoretical connection to interpretation, action, and choice.

This state of affairs is certainly due to the fact that the questions asked by cognitive science and social psychology are very domain-specific, so that theoretical and empirical answers provided by the different research paradigms within the dual-processing tradition do not necessarily combine into a coherent picture. As Smith and DeCoster (2000) point out, many dual-processing accounts one-sidedly emphasize one determinant of the modes over the others, simply because the experimental procedures used to test the particular theories warrant that the neglected variables can be assumed to be "available and unproblematic."[11] Thus, the high domain-specificity of existing dual-process models prevents a more general look at the findings, and their integration into a coherent and general model (Smith & DeCoster 2000, Evans 2008).

The lack of formalization in existing dual-process theories has another drawback: as it is, existing theories often tend to create lists of important determinants ("moderators") without bringing them into a functional relationship (Esser 2001: 257, Mayerl 2009: 13). In particular, this is true for the specification of the interaction between the "match" of symbolically charged situational elements with stored knowledge structures and the other determinants of the processing modes, such as opportunity, motivation, and effort. The current state of dual-processing theory lessens its attractiveness as a main explanatory vehicle for the phenomenon of interpersonal trust. From a methodological standpoint, the lack of formalization of the precise interplay of these variables is a notable flaw, because it is not possible to derive precise and testable hypotheses (Kroneberg 2006a, Mayerl 2009). Essentially, the paradigm does not provide a solid micro theory of action which can establish a causal mirco-mirco transition in our logic of explanation.

Furthermore, dual-process theories, although they certainly admit the idea of context-dependence, do not directly and systematically incorporate the *social* definition of the situation as a conceptual feature, that is, they do not incorporate the fact that the environment which informs perception and choice is always socially prestructured. In the dual-process accounts, the inclusion of structural social conditions is achieved by a translation into motivational and capacity-related constructs. As pointed out, dual-process models pick up the effect of cognitive categorizations (i.e. a sudden "mismatch") only via their indirect effect on other determinants, such as motivation. Likewise, the extent to which a situation is regarded as

[11] "For example, people generally have access to information needed to formulate an attitude about an object whenever they are motivated to do so (Fazio 1990), so little theoretical attention need be given to cognitive capacity. Conversely, participants in problem-solving studies in cognitive laboratories are assumed to be motivated by the task instructions to attempt to perform the task adequately (Sloman 1996), so theories can emphasize capacity and take motivation for granted" (Smith & DeCoster 2000: 125).

structured by social norms ("situational strength") is picked up by current dual-process theories only indirectly through an effect on motivation ("accuracy" and "impression" motivation, Chen & Chaiken 1999, Fazio & Towles-Schwen 1999: 100f., see also Strack & Deutsch 2004), even if the requirement for better theoretical elaboration has been recognized (Kay & Ross 2003, Smith & Semin 2004, Kay et al. 2008). In stark contrast, sociological theories of action have stressed the importance of symbolically structured and *socially* defined situations to action in general (Mead 1967, Blumer 1969) and to the establishment of interpersonal trust in particular (Lewis & Weigert 1985b, Jones & George 1998), which, above all, is a social phenomenon that cannot be explained with exclusive reference to intra-individual (dual-)processes of cognition. Aiming for a broad conceptualization of the social phenomenon of interpersonal trust, it is imperative to include the objective, *social* definition of the situation in the set of central variables of the model, while at the same time keeping up the important notion of adaptive rationality in theorizing about interpretation and choice.

Although important insights can be gained by adopting a dual-processing perspective, it has several downsides that limit its potential use as an explanatory vehicle for the phenomenon of interpersonal trust. From a methodological standpoint, it is a hindrance that a causal link to choice and action cannot be established apart from very general propositions. Focusing on single determinants of information processing, the interplay of the factors has been relatively neglected, so that the notion of adaptive rationality remains somewhat mysterious: how precisely is the degree of rationality connected to opportunity, motivation, accessibility, and effort? How does adaptive rationality translate into action? Furthermore, the social environment is of prime importance in the establishment of interpersonal trust. This goes beyond a mere dual-process notion of person-perception and social cognition (Fiske & Neuberg 1990, Fiske et al. 1999) because the institutional and cultural structure of a trust relation has an influence over and above the cognition of individuals. The sources of familiarity, taken-for-grantedness, and routine which support unconditional trust are found in the structural conditions surrounding the trust relation, and therefore may root unconditional trust in other factors than an automatic application of pre-established relational schemata or stereotypes.

What is needed, taking everything into consideration, is a theory less specific and more general than existing dual-process models, in the sense that it must combine a specification of the processes of interpretation and choice with a direct reference to adaptive rationality. If the causal mechanism behind adaptive rationality can be specified and connected to those related to action, the explanatory power of existing theoretical frameworks of trust could be greatly improved because a reductive explanation of the different "types" (of automatic versus rational, conditional versus unconditional trust, etc.) from a more general theory is possible. To advance our understanding of interpersonal trust, then, is to go beyond the descriptive work of creating typologies and to causally model adaptive rationality. Essentially, we need to specify

a comprehensive micro theory of action which can provide and establish the necessary transitions between structural conditions and aggregate outcomes in our logic of explanation and at the same time accommodate for the impact of adaptive rationality on interpretation and choice.

4.5. The Model of Frame Selection

4.5.1. Modeling Adaptive Rationality

In the following, I want to show how trust can be understood from a perspective of adaptive rationality. To this end, the focus is on the micro-theoretical core on which any explanation of a social phenomenon rests: the general theory of action that is being used. As I have argued, both the rational choice paradigm and the dual-processing approach entail a number of clear disadvantages which lessen their attractiveness as a vehicle of explanation. These shortcomings have been discussed in the previous chapters at large. In my view, the most problematic state of affairs in current trust research is that neither of the available micro theories of action is capable of reflecting what we already know about adaptive rationality, let alone boiling the concepts down into a formally precise and tractable model.

A theory that incorporates both aspects of a social definition of the situation and the idea of human adaptive rationality at the same time is the "Model of Frame Selection" (see Esser 1990, 1991, 2001, Kroneberg 2006a, Esser 2009, 2010, Kroneberg 2011a). The Model of Frame Selection (MFS) assumes that a subjective definition of the situation is a necessary and central condition for establishing and maintaining the capacity to action (Esser 2001: 239f.). The completion of this interpretive process is accompanied by the activation of mental schemata which contain situationally relevant knowledge structures. In the context of interpersonal trust, these may include, for example, specific or generalized expectations, role expectations and social norms, knowledge about institutional mechanisms, and relational schemata linked to particular significant others, such as a familiar trustee. All in all, the totality of activated mental schemata structures the perspective of the actors in a given situation.

It is helpful to analytically separate mental schemata into two broad classes: (1) frames, that is, mental models of typical situations, and (2) scripts, broadly understood as "programs of behavior," that is, mental models of typical sequences of action in typical situations (see chapter 3.1.1 already). Importantly, the concept of a *frame* refers to a socially shared interpretive scheme, or a "situational taxonomy" (Kramer 2006), which actors use to make sense of a given situation, answering the question "What kind of situation is this?" (Goffman 1974). Note that most frames are part of the socially shared stock of knowledge, or "culture" of a society, and they are internalized and learned during socialization (Berger & Luckmann 1966). The activation of a frame defines the primary goals of the social situation, simplifies the individual

goal structure, and prescribes the relevant social production functions (Lindenberg 1989, 1992). The framing of the situation is influential because, once the situation is defined according to a frame, certain programs of behavior, routines, values, and even emotions are activated in the form of associated scripts. In this way, the process of framing and the definition of the situation limit the set of possible and "meaningful" courses of action. Frames direct attention toward specific elements within the situation—they are "selective" and constitute a heuristic which actors use to simplify the process of interpretation. Being part of the actor's stock of learned associative knowledge, frames are connected to situational objects which indicate their appropriateness, that is, they readily contain the heuristic cues which function as significant symbols and trigger their activation.

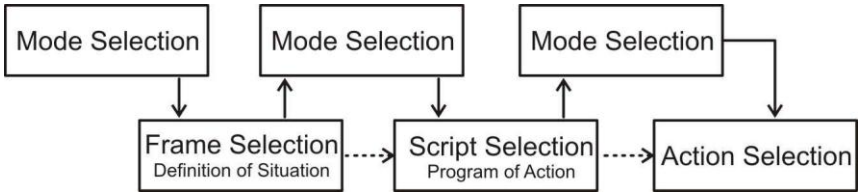
The activation of a frame, according to the principle of spreading activation, activates scripts which have been associated with the frame in the past. *Scripts*, in contrast to frames, have a direct reference to action and choice in that they contain relevant declarative and procedural knowledge, and are organized with respect to goals (Schank & Abelson 1977). They present mental models of typical actions within typical situations, and therefore rely on a preceding successful definition of the situation. The concept of a script will be used broadly in the following to include social norms and roles (Elster 1989) and cultural conventions (Bourdieu 1984), as well as routines, habits, and emotional programs that might become activated in a particular frame and situation (see Abelson 1981). Since scripts are context-specific, the selection of a script can only occur after a frame has been selected, and the situation has been initially defined. Together, frames and scripts provide actors with the knowledge critical for answering the questions, "What kind of situation is this?" and, "How am I expected to behave in such a situation?" The selection of both frames and scripts is guided by the "logic of appropriateness." That is, actors are first and foremost motivated to most accurately decode the situations they encounter in order to decide what to do.

In their combined effect, frames and scripts define the actor's point of view in a particular situation. One important implication of this concerns the preferences of actors, which are variable and influenced by situational circumstances within the MFS. In contrast to the rational choice paradigm, preferences are not treated as fundamental, axiomatic initial conditions, but they are coded in the temporarily accessible mental schemata available to the actors. These may change in response to the context, and so may preferences. In consequence, the scope and application of rational choice models which assume a certain utility function (for example, a social preference such as reciprocity) is necessarily limited, in the sense that they apply only to situations where the actor has already activated a corresponding frame (for example, "situation of social exchange") and a relevant script (for example, "norm of reciprocity"), which jointly generate the preference in question. Since frames and scripts are part of a society's cul-

ture, the actor's preferences are, for the most part, culturally determined (Schütz & Luckmann 1973: 243f., 261f., Fehr & Hoff 2011).

After having selected a frame and a script an actor must finally choose an answer to the question, "What am I going to do?"[12] Note that answers to this question are to a large extent structured by the activated frames and scripts. Activated frames and scripts narrow down the feasible set of alternative actions, they shape the preferences used to evaluate the consequences of action, and they influence the formation of expectations. Thus, when an actor decides on a course of action, his choice has been significantly shaped by preceding interpretive processes and the mental schemata activated in their course. Taking things together, a link from interpretation to action proceeds in three stages which jointly lead to a behavioral response in the form of an action: actors first select a frame with which to interpret a situation (frame selection), they then select a script and program of behavior which is deemed relevant given the frame (script selection), and they finally choose an action (action selection, see figure 16):

Figure 16: The model of frame selection, adapted from Kroneberg (2011a: 128)



As indicated in the figure above, all selections may occur with a variable degree of rationality, that is, with a particular processing mode. In keeping with our terminology, the processing modes will be called the *automatic* and the *rational* mode. The properties of the two modes were described in chapter 4.2.1: the rational mode represents a conscious and deliberative choice, in which the particular alternatives for a selection, their consequences and probabilities, costs and benefits, are analyzed and systematically evaluated. This approximates the maximization of subjective expected utility. In the automatic mode, an alternative is selected unconsciously, based on its temporary accessibility, and following immediate cues in the environment. The MFS holds that each selection (frame, script, action) can occur in an automatic or rational mode.

[12] Decision-making theories such as RCT typically focus on this third and last stage of the sequence. By including the interpretive stages of frame and script selection, the MFS admits that behavior is structured to a large extent by activated mental models, which in turn depend on the properties of the social situation. Therefore, the model can explain the variability of preferences which are taken for granted in economic models.

4.5.2. The Automatic Mode

Before thinking about how the processing mode and the "route" on which trust is built up is determined, we have to define the *selection rules* for each stage (frame, script, action) which govern their activation (see Kroneberg 2006, 2011a: 129ff.). In other words, we ask how the selection of a mental schema can be modeled and formalized, given that a particular mode is in effect. Selection rules specify the causal mechanism that defines the outcome of each stage, and therefore are the "nomological core" of our theory of action. In this sense, they are a most important aspect in an explanatory theory of trust that goes beyond mere descriptive accounts or typologies. Assume that the actor has to "select" alternatives among a set of frames $F = (F_1 ... F_N)$, scripts $S = (S_1 ... S_N)$, and actions $A = (A_1 ... A_N)$. Note that the term "selection" is used here to denote that one mental schema is selected out of the set of potential alternatives, and is then activated. This does not mean that a "selection" is necessarily a conscious choice made by the actor. On the contrary, the automatic mode is characterized by the distinct absence of systematic processing efforts and consciousness.

A selection in the automatic mode is based on the temporary accessibility of mental models. It corresponds to undisturbed pattern recognition and a routine execution of knowledge by the associative cognitive system. Therefore, the selections are based on immediate situational perceptions and the resulting activation level of mentally accessible schemata. As pointed out by Strack & Deutsch (2004), the pathway to behavior in the automatic mode is via a spreading activation of behavioral schemata. Therefore the *activation weights* (AW) of the schemata are decisive in determining which alternative is ultimately selected. The simple selection rule governing all selections in the automatic mode *is to activate the alternative with the highest activation weight*. Thus, based on the spreading activation of associative knowledge initiated by the perception of heuristic cues, actors interpret a situation with the most accessible frame, activate the most accessible script, and execute the most accessible course of action, given a frame and script. This all happens automatically and without a conscious or deliberative effort on the part of the actor. In order to give this proposition substance, we have to define the activation weights of each stage.

A frame selection in the automatic mode is directly guided by the experienced *match* m_i between a frame and the cues available in the situation. In other words, the activation weight of the frame directly corresponds to the fit between stored situational knowledge and the perceived situational cues. According to Esser (2001: 270), the activation weight of a frame depends on three factors:

(1) the *chronic accessibility* a_i of the frame, that is, the actor's general disposition to interpret situations in a specific way. Chronic accessibility of a frame denotes an individual actor's "disposition" to interpret a situation according to stored knowledge which has a high activation potential. It specifies how easily a frame can become activated, and how strongly it is an-

chored in the associative memory system. The parameter is directly related to socialization, experience and learning, and it represents a relatively stable individual property.[13]

(2) the presence of *situational objects* o_i in the situation. These serve as a heuristic cue and are triggered to indicate the validity and appropriateness of a frame. It is through this parameter that the model captures objective-situational variance and the influence of the context on the activation weights. Any element of the situation can become a situational object—items, individuals, actions, or communications. The relevant condition is that the cue is salient and indicative of the applicability of the frame under consideration. This is in turn determined by:

(3) the associative *link* l_i between the frame and the situational object. This individual parameter captures how strongly situational cues are connected to a particular frame and signify and symbolize a certain meaning of the situation to the actor. It thus captures an aspect of "associative strength" (see Fazio 2001, 2007) between an object and its mental representation.

All three factors jointly represent the necessary conditions for a high match. Formally, this implies that the parameters must be related to each other multiplicatively. To formalize the activation weight of a frame, we can hence write:

$$AW(F) = m_i = a_i * l_i * o_i \qquad \text{all parameters } \varepsilon\, [0,1]$$

In the automatic mode, the frame with the best fit to the present situation—that is, the frame with the maximum match—is selected. Generally speaking, the match and activation weights of a frame represent how familiar a situation is to the actor. With a perfect match, the environmental cues can be easily decoded using available knowledge. On the other hand, if the match is low, then the situation is unfamiliar to the actor, and, in the words of Schütz, "routine knowledge not sufficient to master it." New and unfamiliar situations will express themselves through a low chronic accessibility of a relevant interpretive scheme ($a<1$) and/or in a weak associative link between the frame and environmental cues ($l<1$), which then cannot be properly decoded even when they are objectively present ($o=1$). On the other hand, if the actor has strongly anchored the frame ($a=1$) and knows the relevant cues ($l=1$), the match may also be low if the situation is ambiguous and the cues in the environment do not unambiguously indicate a certain meaning (that is, if $o<1$).

[13] Kroneberg uses the term *availability* to describe "general dispositions to adopt a certain interpretation" which are based on "divergent experiences throughout the life-course that vary systematically with socialization in different social contexts" (2011a: 130, present author's translation). However, Higgins (1996) uses the term *accessibility* to describe the activation potential of knowledge, and proposes that *chronic accessibility* indicates "individual differences, including crosscultural differences, in the 'theories' or viewpoints people possess" (1996: 139, see chapter 4.3.3). Comparing the terminology, it is apparent that Kroneberg's "availability" refers to "chronic accessibility" in Higgins' sense—a persisting, long-term individual property that defines the general accessibility of the frame. To minimize confusion, we adopt Higgins' terminology in the following, and differentiate between the aspects of chronic and temporary accessibility.

Likewise, the activation weight of a script depends on several interrelated factors (see Kroneberg 2011a: 131f.). As with the activation weight for a frame, these factors represent the necessary conditions for a high activation weight of a script:

(1) the *chronic accessibility* a_j of the script, denoting how "strongly" the script is rooted in the associative memory system. Similarly to a frame, a script can be more or less chronically accessible and thus feature a lower or higher latent activation potential in the associative memory system. For example, think about a social norm. The degree of norm-internalization can be directly reinterpreted in terms of chronic accessibility. Highly internalized norms will more easily be retrieved as behavioral schemata guiding behavior, because they are chronically accessible to the actor. Similarly, when thinking about routines, chronic accessibility corresponds to the degree of "habitualization" of the routine. All in all, the variable denotes a relatively stable "trait" of the actor.

(2) the *temporary accessibility* $a_{j|i}$ of the script. This variable captures a situational influence in script activation in terms of two sources. First, temporary accessibility depends on the associative strength between the (situationally relevant) frame and a script. Given that the situation has been defined in a certain way, certain knowledge structures, by means of spreading activation, will be more or less accessible.[14] Second, the temporary accessibility of scripts can also be influenced directly by the presence of situational objects indicating the appropriateness of certain actions. In short, temporary accessibility can be altered by internal and external factors.

(3) the match m_i of the activated frame, which is assumed to have an independent influence on the activation of scripts in general. In short, the more familiar a situation is, and the more "certain" the actor is about the validity of a particular interpretation, the higher is the probability that a script related to the frame will be activated. Taken together, the activation weight of a script, including all three as necessary conditions, can be multiplicatively written as:

$$AW(S|F) = m_i * a_j * a_{j|k} = AW(F) * a_j * a_{j|k} \qquad \text{all parameters } \varepsilon\, [0,1]$$

Script selections in the automatic mode follow the rule that the script with the highest activation weight is selected. From the above equation, we can see that this not only depends on the experienced "match," and therefore on the amount of (or lack of) ambiguity with which a situation has been defined. In addition, the frame needs to be associated with a certain script and to "point towards it," and the script needs to be chronically accessible in memory.

[14] For this reason, the accessibility parameter is written with reference to the frame: $a_{j|i.}$ At the same time, the influence of situational objects could in principle be accomplished by breaking the parameter down into additional factors and modeling their interplay (Kroneberg 2011a: 132). This can be done whenever it is of analytical importance. For ease of exposition, we will stick to the sparser notation.

The selection of an action in the automatic mode occurs within a predefined situation, and only after a relevant script has been activated. Routine action is possible only if the script does regulate the course of action to a satisfactory degree: although scripts, as "programs of behavior," contain knowledge about typical sequences of action and expected behavior in typical situations, they are often open to interpretation and may not be detailed enough to unambiguously select one action out of the set of potential alternatives. That is, the activation weight of an action depends on the *degree of regulation* $a_{k|j}$ to which the script dictates a certain action (compare, for example, the famous "restaurant" script, which is relatively open, to a rule-based social norm such as "do not lie!"). If the script does not regulate the course of action to a satisfactorily degree, then routine action is not possible. Second, the activation weights are also dependent on the overall activation level of the script. That is to say, if actors are uncertain about the appropriateness of a script, then a spontaneous and automatic behavioral response based on routine is unlikely, even when the script does regulate the action to a high degree. Thus, the activation weight of an action can be summarized as:

$$AW(A|S,F)) = AW(S|F) * a_{k|j} \qquad \text{all parameters } \varepsilon [0,1]$$

Note that, by including the activation weight of the script, all parameters in the two preceding stages are also relevant necessary conditions for the automatic selection of an action. An actor will automatically select the action with the highest activation weight whenever the situation can be defined unambiguously (a high match m_i), an appropriate script is accessible (a high activation weight AW(S|F) of the script), and the course of action is strongly regulated (high degree of regulation $a_{k|j}$). Typical instances would be, for example, routine everyday behavior and unconditional norm compliance in highly typical situations. On the other hand, even when situations are completely unambiguous and an appropriate script is available (AW(S|F) = 1), activation weights will be low if the script does not regulate action to a degree that allows for a spontaneous execution of stored behavioral schemata.

4.5.3. The Rational Mode

In the rational mode, the selection rules for each alternative and stage are built on a quite different logic. Actors will compare, evaluate, and select available alternatives in an effortful and deliberative reasoning process. In doing so, they also follow a "logic of appropriateness" in that they are motivated to identify the most appropriate alternatives, given the situational circumstances. Since the selections resemble a rational, utility-maximizing choice, the selection rules of SEU theory can be applied to model the stages (see Kroneberg 2011a: 135f.).

Frames and scripts differ with respect to their effectiveness in defining a situation appropriately and in helping to identify the correct course of action. Thus, when elaborating on a proper frame or script, actors are primarily concerned with the question of whether or not the alternative they scrutinize is acceptable. Instead of weighing costs and benefits for each alternative

(as is the case when rationally choosing an action), actors form an expectation about the appropriateness of the considered frame or script. In the case of a frame, the *appropriateness belief* p_i closely corresponds to the match m_i—however, it is in fact perceived and experienced as an expectation. All factors which determine a match in the automatic mode can become problematic and subject to an elaborate reasoning process, in the case of a frame selection in the rational mode: actors think about the fit between situational stimuli and available interpretive schemata, and in doing so, they can contemplate on the presence of relevant situational objects (o_{rc}), the significance of these cues for the particular frame (v_{rc}), as well as the appropriateness of a particular interpretation (a_{rc}) itself. Overall, the appropriateness belief p_i of frame i is forged from these parameters, so that:

$$p_i = o_{rc} * v_{rc} * a_{rc} \qquad \text{all parameters } \varepsilon\, [0,1]$$

As stated before, "appropriateness" is the main concern governing selection of a frame in the rational mode. This means that the motivation behind, and the utility attached to, the different alternative frames, is constant over the range of alternatives. It is, quite generally speaking, that utility which actors derive from forming an appropriate perspective of the world. It can be represented by introducing some constant utility term U_{app}, so that the expected utility of a frame can be written as:

$$SEU(F_i) = p_i * U_{app}$$

In this sense, the decisive factor during a rational and conscious definition of the situation is the appropriateness belief p_i.[15] In a similar fashion, actors can elaborate on the appropriateness of a script in order to answer the question "Which behavior is appropriate in the particular situation?" The *appropriateness belief* p_j of a script can be constructed similarly to the appropriateness belief of a frame. It addresses the questions of whether situational cues are available; whether they are significant for the particular script, and whether the script is appropriate given the frame, so that:

$$SEU(S_j) = p_j * U_{app}$$

Lastly, the selection of an action in the rational mode is characterized by a conscious and elaborate evaluation of the available alternatives in terms of expectations, costs, and benefits. To this end, one can utilize the apparatus of rational choice theory to model the choice of an action. Importantly, the MFS interprets these models *substantially*, that is, as expressing the process of rational choice in a psychological sense (Kroneberg 2011a: 142). For example, in applying SEU theory, one can denote the expected utility of each alternative action as:

[15] See Kroneberg (2011a: 137f.) for some important exceptions, including the case of wishful thinking and the impact of emotions, which can be modeled as additional utility terms.

$$SEU(A_k|F_i, S_j) = \sum p(F,S) * U(F,S)$$

The brackets indicate that both expectations and utility (i.e. preferences) depend on the preceding selection of frame and script. It is apparent that any action selection (both in the rational and automatic mode) is structured by the processes of frame selection and script selection. This pinpoints the importance of the definition of the situation for decision making and the choice of action: frames and scripts activate specific knowledge structures, such as (primary) goals, values, emotions, and programs of behavior. These have a direct effect on expectations and utility. The activation of frames and scripts narrows down the "feasible set" of alternatives that come into question, and which can be scrutinized at all. It also determines how consequences are evaluated. In other words, frame and script activation shapes the preferences and the individual goal structure of an actor.

Having defined the selection weights in the rational mode, let us formally establish the selection rules governing each selection in the rational mode:

$$(1)\ F^* = \text{argmax } SEU(F_i) \text{ for all } F \in F(F_1, \ldots F_n)$$

$$(2)\ S^* = \text{argmax } SEU(S_j|F_i) \text{ for all } S \in S(S_1, \ldots S_n|F_i)$$

$$(3)\ A^* = \text{argmax } SEU(A|F_i,S_j) \text{ for all } A \in A(A_1, \ldots A_n|F_i,S_j)$$

For example, an actor will select that frame F_i for which $SEU(F_i) > SEU(F_j)$ for all $j \varepsilon F$, $j \neq i$ and will select a script for which $SEU(S_i) > SEU(S_j)$ for all $_j \in S$, $j \neq i$, respectively. He will choose that frame, script, and action which, given the alternative set, has the highest expected utility.

The selection rules which govern the selections in either mode establish a causal link between interpretation and choice in both modes. In formalizing these processes, the MFS draws from important insights gained in the dual-processing paradigm, and it also utilizes the framework of rational choice theory. Within dual-processing accounts, the activation levels of mental schemata and their temporary accessibility are regarded as crucial determinants governing automatic activation and use of stored knowledge (Fazio 1990a, Fiske & Taylor 1991, Higgins 1996, Strack & Deutsch 2004). These concepts have been formalized and rendered more precise in the preceding section, by specifying the components of the activation weights and defining their functional relationships. From rational-choice theory, the model adopts the axioms of instrumental rationality to model the selection of frames, scripts, and actions in the rational mode. In doing so, it links rational choice concepts to more fundamental categories found in cognitive science: the accessibility of stored knowledge, the associative strength of mental schemata, and the heuristic cues and situational stimuli of the environment. Importantly, the rational selection of frames and scripts is guided by the formation of appropriateness beliefs

during a conscious elaboration on the presence of situational stimuli and their match with available knowledge. Thus, the process of expectation formation, implicitly assumed in rational choice accounts of trust, is made transparent by specifying the parameters and elements which influence the degree of ambiguity experienced within a situation, with respect to stored mental schemata, that is, by formalizing appropriateness beliefs and by pinning down the cognitive foundation in the form of frames and scripts. At the same time, the rational choice of action is reconstructed as a special case of a more general principle—that of adaptive rationality in interpretation and choice. Thus, both conditional and unconditional decisions, intuitive and intentional choices, deliberate and automatic inferences and interpretation, can be recast in terms of the more general process of adaptive rationality.

4.5.4. The Mode-Selection Threshold

Selection rules establish a link between mental schemata and their activation for each mode. But the MFS additionally tries to explain under which conditions a specific mode will govern a particular selection (see Esser 2001: 257f., Kroneberg 2011a). To begin with, if the degree of rationality involved in interpretation and choice is assumed to be variable, then the mode of information processing itself must be thought of as the outcome of some process—this process is termed *mode selection*; it determines which selection rules the actor applies in a situation. Clearly, mode selection should be governed by the "sufficiency principle" (Chen 1980) and effort-accuracy tradeoffs (Payne et al. 1993): while the actor would always prefer to be most accurate in his selections, the costs and efforts associated with elaborate processing counterbalance the tendency to intervene with an engagement of the rational mode.

Mode selection is an unconscious, preattentive, and autonomous process that determines whether an actor shifts attention to an issue or not; whether an actor does subjectively face a selection problem of interpretation and choice at all—or whether he spontaneously activates the alternative with the highest activation weight without conscious awareness and attention. A model of mode selection must incorporate the insights of dual-processing theory into the determinants of information processing. But to go beyond a listing of potential "moderators," it is necessary to derive a functional relationship between the determinants of opportunity, motivation, accessibility, and effort by formulating their interdependencies in a decision-theoretical framework. Then, by linking mode selections to selection rules, we can establish the link between adaptive rationality, interpretation, and choice.

Trivially, the alternatives which can become an outcome of the mode-selection process are the *automatic* and *rational* modes of information processing. In order to make the process transparent, the MFS formalizes it *in analogy to* a subjectively rational decision. That is to say, although mode selection *does not* represent a conscious maximizing choice, the apparatus of rational choice theory will be utilized in order to systematically derive and formalize the process

in decision-theoretical terms. The "sufficiency principle" is then embodied in the decision-making rules used to model the *optimal* allocation of cognitive resources. For the purpose of modeling adaptive rationality and carving out the "decision logic" behind mode selection, we need to pin down the expected payoffs of both alternatives.

As pointed out before, mode selection is subject to effort-accuracy tradeoffs, the rational route being the more effortful alternative, which potentially (but not necessarily) provides more accurate results. Whether the activation of the rational mode pays off as compared to reliance on the autopilot, is crucially dependent on two factors (Kroneberg 2011a: 145f.): (1) the (non)existence of opportunities to engage in a more elaborate reasoning process and (2) the presence or absence of potential inference errors that an actor commits when actually following the alternative which would be activated in the automatic mode. Jointly, these factors define four states of the world, in which opportunities are (or are not) present and in which the immediately available schema is (or is not) appropriate.

The true state of the world is not known with certainty to the actor, but learning, experience, and situational criteria permit an actor to have a subjective estimate of it. However, these assessments are not part of conscious experience. In sticking with the SEU analogy, we will refer to them as "expectations," but it is important to understand that all the parameters of mode selection cannot be consciously accessed. They represent a result of preattentive environmental scans and, in this sense, they constitute a "natural assessment" (Kahneman & Frederick 2002) achieved autonomously by the cognitive system, as is the selection of the processing mode itself.

Formally, let p (0,1) denote the assessment of sufficient opportunities for reflection. This expectation corresponds to the probability that elaboration in the rational mode is feasible and can be successfully accomplished. Conversely, (1—p) indicates how likely it is that the current situation does *not* afford enough opportunity to engage the rational mode. Second, the activation weights AW(...) of an alternative frame, script, or action are used to assess the probability that the particular alternative is in fact optimal. Thus, a high activation weight indicates that it is appropriate to activate the alternative. Likewise, (1—AW) represents the probability that an inference error is made when selecting the alternative under scrutiny, which potentially incurs some cost. For each selection stage, the corresponding activation weights $AW(F_i)$, $AW(S_j|F_i)$ and $AW(A_k|F_i,S_j)$ will be relevant. The assumption of a direct link between activation weights and inference error assessments shows how immediate perception and spreading activation translate into processing-mode determinants.

Since the mode selections in each stage are formally identical, we restrict ourselves in the following to presenting the determination of the processing mode for a frame. Similar formulations for the stages of script and action selection can be derived by replacing the correspond-

ing activation weights. In the case of a frame, the relevant activation weight is the match m_i, which indicates the fit between the frame i and the current situation. Upon entering a situation, some *initial* frame i will attain the highest activation weight, and will therefore be subject to the question of whether or not it should be followed automatically ("initial categorization," Fiske & Neuberg 1990). Note that an alternative frame will be considered only if the appropriateness of the initial frame is doubted, and thus if frame selection occurs in the rational mode. We can interpret the match m_i as the actor's expectation that frame i is in fact appropriate. Conversely, (1—m_i) represents the probability that frame i is not appropriate, and some other interpretive schema is correct.

Combining both elements, we can construct inferences of the probabilities of the four possible states of the world. These represent natural assessments of whether (or not) opportunities for reflection are sufficient and whether (or not) the initial frame is appropriate. For example, the probability of the occurrence of a state of the world in which no opportunities for reflection exist, and in which the initial frame is valid is (1–p) * m_i. Likewise, p * (1-m_i) corresponds to the probability that sufficient opportunities exist and the initial frame is not valid. Having defined subjective probabilities thus, we are left to define the actual payoffs to a selection made in each mode and under each *true* state of the world, in order to model the consequences that the selections in different modes have in each state of the world (table 2):

Table 2: Mode-selection and the subjective states of the world

Alternative States of the World and Subjective Probabilities of Occurrence				
	(1)	(2)	(3)	(4)
Opportunity sufficient?	Yes	Yes	No	No
Initial Frame Valid?	Yes	No	Yes	No
Subjective Probability	p * m_i	p * (1-m_i)	(1-p) * m_i	(1-p)* (1-m_i)
Mode Selection Outcome:				
Rational Mode	U_i - C	U_{rc} - C	U_i - C	-C_f - C
Automatic Mode	U_i	-C_f	U_i	-C_f

The utility associated with the initial frame i is U_i. This utility can be realized whenever frame i is selected and i is in fact the true state of the world (cases 1 and 3). The adoption of frame i in a state of the world in which it is *not* valid results in a wrong definition of the situation, in which case the actor incurs some costs C_f>0 if he follows his initial categorization (cases 2 and 4). The engagement of the rational mode is always associated with costs C>0, representing the mental effort incurred in the form of time and energy consumption. The actor can fruitfully "capitalize" on these costs if sufficient opportunities do exist and if the initial frame i is *not* valid (case 2). In this case, an alternative frame j will be discovered in the process of forming appropriateness beliefs, and the actor can realize some alternative utility U_{rc} associated with the adoption of this frame j.

In all other cases, the actor would be better off following the initial frame in the automatic mode: If frame i is valid and opportunities exist, then the actor will rationally discover that the initial frame is valid, but also incur costs which could have been saved (case 1). If frame i is valid and opportunities do in fact *not* exist, then the reasoning process will not lead to a successful redefinition of the situation, and the actor will have to rely on his initial frame i, while incurring the processing costs (case 3). If frame i is *not* valid and opportunities do *not* exist, then the actor will not discover the appropriate alternative, but will make a false interpretation and incur the costs of rational processing C on top of it (case 4).

We can derive the expected utility associated with the two modes by following the principles of SEU theory. Weighing the payoffs of each consequence with their associated probabilities, we get:

$$SEU(automatic) = p*m_i*U_i + p*(1-m_i)*(-C_f) + (1-p)*m_i*Ui + (1-p)*(1-m_i)*(-C_f)$$

$$SEU(rational) = p*m_i*(U_i-C) + p*(1-m_i)*(U_{rc}-C) + (1-p)*m_i*(U_i-C)+ (1-p)*(1-m)*(-C_f-C)$$

The selection of the rational mode is contingent on the condition that SEU(rational) > SEU(automatic). Simplifying this inequality yields the following *threshold condition* which governs the activation of the rational mode:

$$p*(1-m_i)*(U_{rc}+C_f) > C$$

This inequality can be interpreted intuitively: actors will engage in an elaborated reasoning process whenever the benefits of doing so outweigh the costs. Thus, the MFS allows us to derive the argument of trading accuracy against effort in selecting decision strategies in a very general fashion, starting from the idea that inferences about the states of the world determine the processing mode. Adding assumptions concerning the possible outcomes in each state of the world, inference errors and their potential costs and benefits, the SEU principles provide an answer that is surprisingly consistent with dual-process research: whenever sufficient opportunity exists (p) and an alternative frame is valid ($1-m_i$), the actor can realize the utility from an appropriate definition of the situation (U_{rc}) and avoid the cost of defining the situation inappropriately (C_f).

The cost-term C, that is, the amount of expected mental effort involved in the rational mode, has both a situational and an individual-intrinsic component. It varies with the complexity of the task (Payne et al. 1993) and with individual processing abilities, such as general intelligence and task-specific skills (Mulder 1986). The term ($U_{rc}+C_f$) represents the opportunity cost of making a false selection (Kroneberg 2011a: 148). It can be easily translated into motivational constructs such as “accuracy motivation” (Chaiken 1980) or “fear of invalidity” (Fazio 1990) and represents the element of motivation as a core determinant of information

processing.[16] In particular, the importance of a selection, which may vary with structural parameters such as stake size, can be captured by C_f. In scrutinizing the model, it is easy to locate the other determinants as well. The element of opportunity is directly introduced with the parameter p, which captures situational constraints such as time-pressure and cognitive load, as well as individual-intrinsic factors such as processing capacity and ability. By treating the activation weights as the relevant informational cues to the generation of subjective expectations about the states of the world, the model directly makes use of schema accessibility, and also establishes a link with the presence of heuristic cues within the environment. It supports a spreading-activation argument in the form of associative links l_i between situational objects and particular mental schemata, and it invokes the hierarchical structure of schemata, spreading activation among frames, scripts, and actions (as formalized in the parameters $a_{j|I}$ and $a_{k|j}$).

We can rearrange the threshold condition to display SEU(automatic)>SEU(rational), so that the model demonstrates the conditions that need to be fulfilled in order to select the automatic mode during frame, script, and action selection:

Automatic Frame Selection:	m_i	$> 1 - C / (p * U)$
Automatic Script Selection:	$m_i * a_{j\|i} * a_j$	$> 1 - C / (p * U)$
Automatic Action Selection:	$m_i * a_{j\|i} * a_j * a_{k\|j}$	$> 1 - C / (p * U)$

In this formulation, it is easy to see that a high match—that is, a clear and unambiguous definition of the situation—is a fundamental precondition for automatic selection in all stages from interpretation to choice. The rational mode is *not* selected if the initial match is high, and high processing costs (C), insufficient opportunity (low p), or a lack of motivation (low U) shift the threshold to a low level. On the other hand, an ambiguous definition of the situation and a low match foster activation of the rational mode. The model restates the default-interventionist proposition that routine persists as long as the environment presents itself as unproblematic. Consider the case where m = 1 and a "perfect match" prevails. Under these circumstances, an actor will always select a frame in the automatic mode.[17] Interpretation is spontaneous and fully automatic, because an appropriate frame is accessible and fits the available situational cues.

However, the automatic activation of scripts and the unconditional execution of actions, in comparison to a frame selection, rest on increasingly stricter preconditions, since additional constraints must be met in order to select the automatic mode. For a script to be selected automatically, it must additionally be stored in the memory system (a_j) and associated with the

[16] We will subsume both elements and use the shortcut notation $(U_{rc}+C_f) = U$ in the following.
[17] Since C >0, U > 0 and p (0,1), the right-hand side of the equation is restricted to the interval $)-\infty,1)$.

particular frame ($a_{j|i}$). If actions are to be selected automatically, the script additionally needs to regulate a course of action to a satisfactory degree ($a_{k|j}$). Only in this case does a direct link between perception and behavior in the sense of a "spontaneous flowing from individuals' definition of the event" (Fazio 1990: 91) to a behavioral outcome exist, which then solely rests on the spreading activation of behavioral schemata (Strack & Deutsch 2004). Such conditions are met, in particular, in the case of social norms, value-based rule systems, and routine habits.

The model demonstrates adaptive responses in the degree of rationality to potential inference errors, as stipulated by Fehr and Tyran (2008). It recasts the idea that inference errors are a primary determinant of managing social situations, as exemplified in the social exchange heuristic (Yamagishi et al. 2008), and as postulated in "error management theory" (EMT, Haselton & Nettle 2006). According to EMT, the relative magnitude of potential inference errors determines the evolutionary direction of prosocial perception biases. However, they are not regarded as a primary key to interpretation. In the "social exchange heuristic" stipulated by Yamagishi et al. (2008), which was introduced at the beginning of chapter 4, actors are assumed to act fully rationally once the automatic process of SEH activation is terminated. However, the MFS suggests that the processing of inference errors occurs on a much more dynamic and continuous basis. The activation of the SEH can be reconstructed as a special problem of frame selection with respect to defining a situation as one of social exchange and activating a corresponding frame and script. In the model of Yamagishi et al. (2008), actors are severely limited with respect to the set of interpretational schemata they possess. The model considers only two frames: "detection / sanctioning" and "no detection / no sanctioning," and actors—somehow—define the situation automatically. Once the situation is defined, actors act fully rationally. However, in the MFS framework, other possibilities exist. Both interpretation and choice can be executed with a variable degree of rationality. This degree may change in response to situational circumstances and potential inference errors that are inferred according to the natural assessments of opportunity and appropriateness. In fact, the processing of fictive error signals and potential inference errors seems to play a key role in trust problems as well, as shown in a fMRI-study conducted by Lohrenz et al. (2007).

The model helps to explain the unconditionally of normative routine often observed in human action, which contradicts standard rational choice theory (see Elster 1989, Boudon 2003). In contrast to a rational-choice explanation, unconditional norm compliance can be understood and explained in terms of automatic schema-activation and routine execution of scripts, addressing a completely different "logic" of action selection than that of the instrumental maximization of subjective expected utility. This is reminiscent of the idea that highly internalized norms can "override" rational choice and that rule-based decisions must be understood as alternatives to consequence-based maximizations (March & Olsen 1989, Vanberg 1994). On the

other hand, if scripts do not regulate action to a satisfactorily degree, actions will have to be selected in the rational mode. In this case, actors must perform maximizing operations of the sort proposed in rational choice models. For example, game-theoretic formalizations are immediately applicable if there is some strategic interdependence: *given that* a particular frame has been selected, a particular script activated, and *given that* action is selected in the rational mode, norm compliance, for example, can be explained in terms of psychological cost-benefit considerations, in which the actors take into account the (dis-)utility stemming from norm compliance, and (potentially) the other-regarding preferences they and others have activated during interpretation.

By specifying the mechanism of mode selection, the model goes beyond the traditional rational choice framework, because the conditions and range of rational choice models and the applicability of economic models in general are spelled out. The introduction of frame selection and script selection as processes prior to action explicates the proposition of "frames moving beliefs moving choice" (Dufwenberg et al. 2011) and of "culture shaping preferences" (Fehr & Hoff 2011). These propositions put forward the importance of a socialized stock of cultural knowledge of typical situations and typical sequences of actions for individual interpretation and choice. Within the MFS framework, we reconstruct a selection that approximates a rational choice merely as a special case, one that actors will perform when routine mental schemata are inaccessible, when situations are interpreted as important and nonroutine, or when the motivation to override automatic categorizations is very high. In the case of automatic selections, the decision logic behind selection of frame, script, and action is completely different to that of preference-based utility maximization, building on the spreading activation of mental schemata which can trigger a direct perception-behavior link.

Note that we can also derive very specific interaction hypotheses from the model which go beyond traditional sociological theories of action, the rational-choice framework, and modern dual-processing accounts. Generally speaking, *all* model parameters are associated with each other and linked to each other, thereby *jointly* influencing the mode-selection threshold. That is, the postulation of interaction effects between the model parameters is a model-inherent feature which always has to be accounted for, and which can be predicted in its direction and scope (see Kroneberg 2011a: 151f., 2011b, and chapter 6 below). A most important implication of the model concerns the maintenance of unconditional routine, and its interplay with the cost-benefit structure of a situation. More specifically, the adaptive rationality approach developed here delivers specific answers to the question of how "high stakes," impact the processing mode and the decisions of an actor, and how incentives interplay with other parameters, such as script internalization and chronic frame accessibility.

According to the economic "low-cost hypothesis" (Diekmann & Preisendörfer 1992, 2003, Rauhut & Krumpal 2008), norm compliance is dependent on the cost and benefits associated

with its implementation—the probability of norm compliance decreases with increasing costs. This is in fact a very general feature of all traditional and broad RCT approaches: norms are part of the cost-benefit calculations that rational actors carry out. Thus, there is always a "price" to norm compliance which, if too high, will be outweighed by the prospective gains of not following the norm (or attitude). According to the economic low-cost hypothesis, attitude-conforming behavior can be expected in low-cost situations only. Looking at the mode selection threshold, we can see that a high match (the match m_i approaches the value of one) can trigger the automatic mode *even if* the direct cost of doing so becomes very high and has severe utility-related consequences. This effect is even more pronounced if no opportunities exist or if the decision process has high cognitive costs C in addition. The special role of the match and the categorizations delivered by the associative memory system in restricting and diminishing the influence of instrumental concerns is a feature which also contradicts the classical "Theory of Reasoned Action" (Ajzen & Fishbein 1980) and the "Theory of Planned Behavior" (Ajzen 1985), it is not predicted in generic dual-process models (Mayerl 2010).

The conceptualization of the match itself reveals another important difference. The magnitude of the match m_i depends on the chronic accessibility of a mental schema (a_i), its link to situational objects (l_i) and the presence of situational objects (o_i). Together, these are *necessary conditions* for a high match, and we can predict an interaction between them: a high match (and the selection of the automatic mode) relies both on the presence of situational cues *as well as* a high chronic accessibility of related mental schemata and their mental association. In contrast, the influence of situational cues is expected to *de*crease in generic dual-process accounts with *in*creasing chronic accessibility of an attitude (Mayerl 2010: 42). All in all, the model presents a range of theoretically new and contrasting hypotheses with respect to the cognition-behavior link to trust which have not been developed in, and are not covered by, existing dual-process theories and the rational choice framework.

4.6. Explaining Conditional and Unconditional Trust

From the perspective of adaptive rationality, the prominent conceptualizations of interpersonal trust found in the literature appear to be much less conflicting than they seem *prima facie*. One fundamental ingredient in a broadened understanding of the phenomenon is the assumption of adaptive rationality on the part of the trustor and trustee, which must be respected and explained as an endogenous parameter during trust development. Research across many disciplines has demonstrated that human rationality is not only bounded, but also flexible and highly adaptive. If we want to advance our understanding of the phenomenon of trust, we have to take these insights to the heart of our theory, that is, to the microlevel theory of action, to the actor models we apply—to trustor and trustee—and to the decision-logic and micro-mirco-

transition that follows. If we take the notion of adaptive rationality to be central to human cognition, then all aspects of interaction must be regarded from that perspective.

If we think about the different "types" of trust that have been put forth by trust researchers from the perspective of adaptive rationality, it is obvious that the most important difference between them is the degree of rationality involved (e.g. calculus-based; conditional trust *versus* affect-based, institution-based, or rule-based trust; and so forth). In essence, *adaptive rationality is a central dimension of the trust concept*. We cannot understand trust without reference to adaptive rationality and individual processing states, and without understanding the underlying mechanisms and processes governing their selection. The framework that will be developed in the following can be used to derive and explain both unconditional forms of trust (trust without doubtful and conscious elaboration), and conditional forms of trust (in which the trustor subjectively faces the trust problem and elaborates on his future course of action). More importantly, the framework we will develop aims for a specification of the conditions that must be met in order for the one or the other type of trust to emerge.

The MFS, by providing a very general perspective on rationality and decision making, allows trust researchers to reinterpret seemingly contradictory and disconnected concepts under the common umbrella of adaptive rationality. To see how we can join existing theory and remedy theoretical problems, it is instructive to review Luhmann's analytical separation of trust, familiarity, and confidence and to contrast his approach to the social-psychological perspective. All of them being "different modes of asserting expectations—different types, as it were, of self-assurance" (Luhmann 1988: 99), the three concepts proposed by Luhmann describe different ways of dealing with ambiguity, a state in which actors regularly remain incapable of action. Trust, according to Luhmann, manifests itself in a "particular style of attitude" (ibid. 27), but, in contrast to confidence, requires a conscious acceptance of risk. For this reason, Luhmann is often pushed into the corner of rational choice and appropriated by rational choice advocates (e.g. Hardin 1993, Sztompka 1999: 60). However, Luhmann is unequivocal in that he does *not* address a rational choice of action with his statements. As he states in the last chapter of his book:

> "If one were to take as a yardstick the concept of rationality in decision-making theories—be it that of the rational choice in the employment of means, or that of optimizing—one would from the outset fall into a too narrow conceptual frame of reference which cannot do justice to the facts of trust. Trust is not a means that can be chosen for particular ends, much less an end/means structure capable of being optimized. Nor is trust a prediction, the correctness of which could be measured when the predicted event occurs and after some experience reduced to a probability value [...] Trust is, however, something other than a reasonable assumption on which to decide correctly, and for this reason models for calculating correct decisions miss the point of the question of trust" (Luhmann 1988: 88).

Luhmann is *uniquely* concerned with the problem of interpretation and the definition of the situation. Recasting his ideas within the MFS, we see that his concept of confidence equates to *frame selection in the automatic mode*, while his concept of trust describes *frame selection in the rational mode*. Quite generally, we can hypothesize that actors will use a larger share of

their cognitive resources when defining situations in the rational mode, accompanied by a selective focus of attention. The trustor then consciously perceives the trust problem; he knows that the result depends on the actions of the trustee and that a failure of trust is among the trustee's viable options. In the case of trustful action, the trustor nevertheless defines the situation sufficiently confidently as to reassure himself about the reasonability of his trusting choice. But with the concept of adaptive rationality at hand, we see that Luhmann's conceptualization is limited and incomplete. His analytical separation of confidence from trust categorically excludes any notion of "unconditional" trust, which other researchers claim to be a primary characteristic of the phenomenon. It also warrants the question of what the difference between confidence and familiarity really is (Endress 2001).

In contrast, recent social-psychological work has objected to Luhmann's position, and linked trust to the routine use of simple inference rules, while claiming that distrust (not trust!) entails the nonroutine mode of information processing, a deliberate assessment of expectations of trustworthiness, and of the intentions of the trustee (Schul et al 2008). Trust and distrust are, in essence, conceptualized as endpoints on a continuum of information-processing states of the cognitive system which are linked to particular subjective experience of the trust problem:

> "When a state of trust is active, one tends to believe, to follow the immediate implications of the given information. In contrast, when a state of distrust is active, one tends to search for non-obvious alternative interpretations of the given information, because distrust is associated with concealment of truth. Thus, in distrust, the mental system becomes more open to the possibility that the ordinary schema typically used to interpret the ongoing situation may need to be adjusted" (Schul et al. 2008: 2).

In this perspective, interpretation and the use of inference rules, as well as the behavioral response, are also *fixed* to the state of information processing, but in a way completely opposite to that proposed by Luhmann. When approximating a favorable and unambiguous definition of the situation, the spontaneous use of schemata becomes more likely.[18] Empirically, Krueger et al. (2007) also show that different neuronal systems become active in trust decisions, depending on the processing strategy which trustors use to make the choice: "Conditional trust assumes that one's partner is self-interested and estimates the expected value of past decisions; ... it is cognitively more costly to maintain. In contrast, unconditional trust assumes that one's partner is trustworthy and updates the value of one's partner with respect to their characteristics and past performance, ... it is cognitively less costly to maintain" (ibid. 1).

[18] "Psychologically, this means that in the former case [of full trust], individuals believe that the other has only benign intentions, shares their interest totally, and what he or she says is unquestionably valid. In the case of extreme distrust, individuals are equally confident that the other's intentions are totally malign, his or her interests are wholly incompatible with their won, and what he or she asserts is best interpreted according to a theory they have about how, given the situation, others would likely try to dupe them" (Schul et al. 2008: 9).

The notion of adaptive rationality in interpretation and choice helps to reconcile such contradictory positions, because it suggests that interpretation, choice, and the degree of rationality involved in either stage must be treated as analytically separate and distinct, yet at the same time flexible. In short, actors can flexibly use different "routes to trust." With respect to the problem of interpersonal trust, the subjective definition of the situation is central. Interpretation precedes the formation and stabilization of expectations of trustworthiness, and it affords a restriction of the feasible set of alternatives by activating trust-related mental schemata. This enables a significant reduction of social complexity. But this process can be automatic or rational, conscious or unconscious—trustors need not necessarily be aware of the trust problem and the relevant expectations of trustworthiness. In concluding interpretation, they have nevertheless acquired a particular attitude towards the situation which structures their perspective. The more problematic the definition of the situation has been, the more likely it is that the automatic process of pattern recognition has been truncated. Conditional and unconditional behavioral trusting strategies can be used by trustors in the aftermath of interpretation. The mode of action selection then depends on how unambiguously the situation could be defined, and how strongly the activated scripts and other schemata regulate action.

A conceptualization of trust along the dimension of adaptive rationality, paired with an analytic separation of interpretation from choice, exemplifies how trust can be understood as a mechanism for the reduction of social complexity, and it gives meaning and substance to the notion of "suspension" and the "leap of faith." If we scrutinize the model to answer the question of where complexity is effectively reduced, the process of frame selection and the definition of the situation naturally acquire the most important role (see Möllering 2006a, b). Framing processes are directly connected to the formation of expectations of trustworthiness, to a shaping of the "initial beliefs," preferences, and the activation of trust-related schemata, as well as emotional programs and values which orient trustful behavior, once they become activated. But all this can happen spontaneously and without any allocation of attention. It can likewise occur with a conscious capture of attention, and in controlled elaboration.

Situations can be defined right from the outset in such a way that the trustworthiness of the trustee is subjectively never questioned. A prominent example would be dyadically embedded exchange relationships, for example a friendship or a partnership. Models of interpersonal trust development (see chapter 3.1.4) can be abstractly understood as the emergence and formation of shared relational schemata which serve as a trust frame for the relationship. They include attributes of trust and trustworthiness and organize the particular trust relation (see chapter 5 below). Similarly to an attitude towards others, trust, in this sense, must be learned and rooted in the memory system in the form of trust-related knowledge. This argument also applies to forms of institution-based trust in which a trustee receives a favorable attribution of trustworthiness in that the trustor can apply a learned categorization (a stereotype, a social

role) or because the interaction takes place in a context where norms and institutions serve as "rounding framework of trust" (Giddens 1990) in that their relevance and structural assurance is recognized and met with a sufficiently strong chronic accessibility of corresponding scripts. Another instance would be everyday routines that involve trust problems which have been previously solved, and now belong to the world of the typical. That is to say, "suspension" and the "leap of faith" occur at the interpretive stage of frame and script selection, and they must be principally understood as a synonym for the automatic activation of relevant trust-related knowledge. We "leap" into trust and "suspend" uncertainty when a schema that harbors sufficiently confident expectations is automatically activated during interpretation. Adaptive rationality and context-dependent mode selections are the key mechanisms behind the "leap of faith."

Focusing on the selection of actions, we can use the notion of "unconditional trust" even with reference to action and observable behavior. That is, the choice of a trusting act can also be automatic if action selection occurs following the activation of a sufficiently regulative script (for example a rule, role, or routine). The mode selection thresholds indicate that this case is tied to more stringent preconditions concerning the internalization of mental schemata, their temporary accessibility, and the degree of regulation of relevant scripts. In short, the complete causal chain of ideal-type "blind" and unconditional trust is located in the context-dependent mode selections from frame and script to action selection; it extends from interpretation to choice. In this sense, *unconditional* trust does in fact equate to a reduction, or "suspension," of risk and ambiguity into subjective certainty, by means of "overdrawing" information as a result of mode selection and automatic schema application (see chapter 2.2.3).

On the other hand, we can also think of trust in terms of the rational and analytic processing of information, and we need to separately address the stages of interpretation and choice again. For one, if the situation cannot be defined automatically, the trustor has to engage in a more elaborate process of interpretation. This has several consequences: first, the trustor necessarily becomes conscious of the trust problem. The failure of pattern recognition will be experienced in terms of low processing fluency and the "doubtful" intervention of the rational system. In scrutinizing the situation, the fact that the choice of a trusting act involves a risk and entails vulnerabilities will enter the subjective experience of the trustor. Second, given that the rational system intervenes, the formation of appropriateness beliefs can be directly related to the generation of trustworthiness expectations, because "appropriateness," in the context of the trust problem, concerns the question of whether or not the situation can be appropriately defined as one in which trustworthiness is warranted. In doing so, the trustor compares and evaluates different potential interpretations which make use of relevant trust-related knowledge, in order to assess the trustee's trustworthiness. Additionally, rational elaboration of the trust problem entails that the trustor, during interaction, acquires additional individuating infor-

mation about the trustee, and performs assessments of trustee characteristics in order to scrutinize his trust-warranting properties of benevolence, integrity, ability, and predictability.

To this end, the trustor accesses and tests a range of available schemata (i.e. stereotypes, roles) for their appropriateness, retrieves individuating trust-related information, and scrutinizes the situation for the "situational strength" of normative regulation, for structural assurance, situational normality, and so on. Of course, he may still arrive at a conclusion that trustworthiness expectations, as suggested by the immediate categorization, are still valid and applicable. But schema-driven categorizations can be "overridden" by the intervention of the rational system. Adopting a very general view on the phenomenon, we cannot, in advance, determine which particular category of trust-related knowledge will become relevant in a specific situation—all we can say is that their activation is guided by a controlled reasoning process, and that, subjectively, the trustor's inference will accumulate and feed into his "feeling of rightness" with respect to the choice of a trusting act. Trust then is "bothersome" (Messick & Kramer 2001), and may be accompanied by anxiety, deference, and doubt. In this sense, *conditional* trust does in fact equate to a reduction of ambiguity into risk by assessing appropriate knowledge (see chapter 2.2.3).

Once the situation is defined, the trustor must make a decision. In the ideal-type case of conditional trust, this also happens in the rational mode of information processing. Economic models of trust and trustworthiness can be used for explaining the choice of a trusting act whenever the elaborated mode of information processing prevails in the final stage of action selection. The trustor will then consciously weigh costs and benefits according to his expectations, and will finally make his decision on the choice of a trusting act. If the context of the trust problem indicates the relevance of social norms, then social preference models can be utilized to explain the choice of a trusting act, meaning that norms are treated as additional arguments in the extended calculations executed by trustor and trustee.

Note that, as the stages of interpretation and choice are principally independent, a rational definition of the situation need not accumulate into the rational choice of the trusting act, nor *vice versa*. Imagine, for example, that the situation is initially ambiguous simply because of "noise" with respect to the situational cues. During a rational interpretation, the trustor can reduce the noise and filter the relevant cues which allow for the activation of some appropriate frame. Given that the selected frame has a strongly regulating script attached to it, the choice of a trusting act may nevertheless be performed automatically, even when the situation has initially been defined in the rational mode. Similarly, a situation that is unambiguous with respect to interpretation may not regulate a trustful course of action to a satisfying degree (i.e. low structural assurance, an open script). Then the choice of a trusting act will be conditional, but the expectations used by the trustor during his rational choice have been preconsciously structured by automatic interpretive processes.

From this dualistic perspective, it is easy to follow critiques of the rational choice approach to trust, which often condemn a systematic neglect of the full range of the phenomenon (Lewis & Weigert 1985b, Williamson 1993, Endress 2002, Möllering 2006b). Instrumental cost-benefit considerations with respect to the consequence of action must be understood as being merely one special case. Conceptually, the term "trust" can be used in two very different ways. On the one hand, it can refer to the structuring of perception and the definition of the situation. In this reading, "trust" stands for the formation and stabilization of favorable expectations embedded in stored mental schemata which contain trust-related knowledge. But we can use the term with reference to the aspect of action and choice, more specifically, with respect to the choice of a trusting act. In this case, the trustor has defined the situation already, and is concerned with the selection of script and action.[19] In acknowledging adaptive rationality, we must conclude that *either* step can occur automatically *or* rationally.

Trust researchers have difficulties in integrating their theoretical contributions precisely because the analysis is usually limited to only one stage (interpretation *or* choice), assuming one mode of information processing (automatic *or* rational), and fully neglecting the potential adaptivity thereof. Trustors, if the definition of the situation is automatic, do not perceive the trust problem as such, because the situation can be *confidently* defined according to existing knowledge structures which include relevant favorable expectations. In the ideal-type case, even the selection of action and the choice of a trusting act occur automatically. This is the idea of "blind" trust warranted on routine grounds that suppresses vulnerability and suspends doubt from subjective perception, and is not accessible to the actor—only a failure of trust will trigger the trustor's realization that a trust problem did indeed exist.

Taken together, in adopting the MFS perspective of adaptive rationality paired with the analytical separation of interpretation from choice, trust research is equipped with a theoretical framework that can be used for the explanation of a broad range of phenomena related to observable trusting choices and trustworthy responses. This necessitates a rethinking and redevelopment of the terminology of trust. "Automatic" and "rational" processes may co-occur within the trust problem, and when theorizing about trust, we have to be clear as to whether we address interpretation *or* choice, rational *or* automatic modes of information processing. The perspective of trust and adaptive rationality developed here also warrants that trust is highly dynamic (this is issue will be further explored in chapter 5).

[19] Hardin's explication of the trust-action relation is a prime example: "Trust is not a risk or a gamble. It is, of course, risky to put myself in a position to be harmed or benefited by another. But I do not calculate the risk and then additionally decide to trust you; my estimation of the risk is my degree of trust in you. Again, I do not typically choose to trust and therefore act. Rather, I do trust and therefore choose to act" (Hardin 1993: 516). His statement indicates that trust is "already in place" when a choice is made—the process of interpretation is implicitly taken for granted.

Generally speaking, we are concerned with the selection of a *trust frame* F_t, which the trustor can use to *unambiguously* define a trust problem. A trust frame entails sufficiently favorable expectations of trustworthiness. As will be argued in the next chapter, relational schemata, by means of which actors frame their personal social relationships, are a particularly important type of trust frame for interpersonal trust. The framing of the situation with the help of a trust frame therefore entails, in the ideal-type case, a subjectively unambiguous and favorable expectation of trustworthiness, which may (or may not) be part of subjective experience. If the trust frame is selected in the rational mode, the trustor perceives the trustee as trustworthy, because his favorable expectations exceed the subjective threshold necessary for inducing trust. Thus, concerning a trust frame F_i, we assume that:

(1) A trust frame is linked to an appropriate script, that is, $a_{j|k}=1$. (A1)
(2) The script regulates action to a high degree, such that $a_{k|j}=1$. (A2)

Since the trustor can activate a script that sufficiently regulates action (for example, the reciprocity norm, rules of friendship, an appropriate social role), there is no doubt as to what the appropriate course of action is—namely, trusting act and trustworthy response—should the trustor start to consciously elaborate on the trust problem in the rational mode. A trust frame, in the ideal case, contains knowledge that suggests a favorable trustworthy response *by definition*.

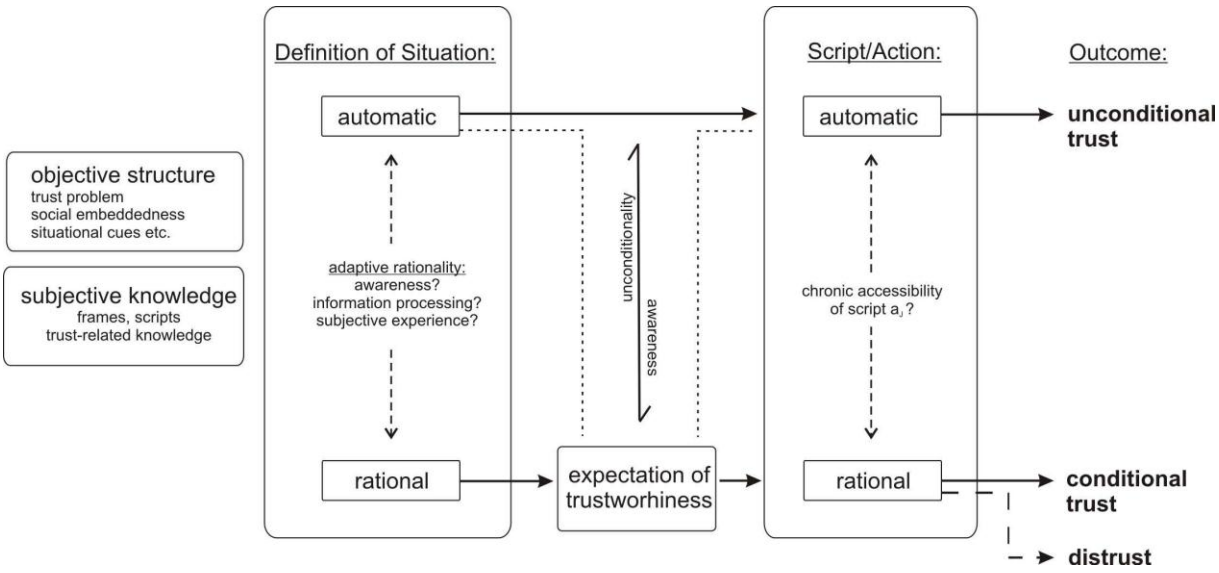
In a given trust problem, the success of pattern recognition and the degree of ambiguity experienced is represented by the match m_t. Thus, when thinking about the activation of a trust frame, we can write the thresholds for the automatic mode for the process of frame and action selection, using A1 and A2, as:

(1) Frame Selection: $m_t > 1 - C / p * (U_{rc}+C_f)$
(2) Action Selection: $m_t * a_j > 1 - C / p * (U_{rc}+C_f)$

The unconditional choice of a trusting act occurs whenever $m_i * a_j > 1 - C / p*(U_{rc}+C_f)$. If the threshold for the automatic selection of a trust frame is satisfied, interpretation, and choice are automatic. This demonstrates the "leap of faith" in trust, as occurring *on the level of mode selection*. The trust problem itself does not get a "grip on the mind" of the trustor, and neither does his expectation of trustworthiness. On the other hand, rational elaboration during interpretation will foster a conscious perception of the trust problem. In the case of rational action selection, trustors will engage in cost-benefit considerations similar to those proposed in economic theory, respecting, for example, both the direct and indirect costs of norm compliance and defection. This additional cognitive effort can be saved if trust is automatic, in which case the trustor chooses a trusting act without any further scrutiny of the trust problem. Both possibilities lead to the observable behavioral outcome: the choice of a trusting act. The trustor can

also choose to distrust. However, this entails an activation of the rational system, harboring doubt, and a breakdown in the routine of pattern recognition. Distrust is always "reflective," in that the trustor will, at least for a minimum of time, pay attention to the fact that a trust problem exists, even if distrust occurs swiftly, or even "automatically", for example, because a relevant stereotype has been activated. The following picture schematically displays the complete model of trust and adaptive rationality as specified in this section (figure 17):

Figure 17: The model of trust and adaptive rationality



In most general terms, we can define trust as *an actor's definition of the situation that involves the activation of mental schemata sufficient for the generation of a favorable expectation of trustworthiness and the subsequent conditional or unconditional choice of a trusting act*. This definition is, of course, very general and does not take care of the respective content of trust-related knowledge. The "typological" specification of trust depends on what category of trust-related knowledge is being used, and in which mode of information processing it is applied. Note that the present definition merges psychological aspects and a notion of trust as a "state of mind" with the behavioral aspect of action. It includes the aspect of intentionality, once we keep in mind that expectations are a precursor to intentions—however, the "willingness to be vulnerable" and the inherent intentionality rise to conscious perception only in the rational mode. The choice of a trusting act can causally be traced back to an attempt at rational inference, at assessing trustee characteristics, and rationally weighing the expected costs and benefits of action and the activation of specific expectations, as well as to a routine execution of trust-related knowledge (relational schemata, rules, roles, routines) and a reliance on heuristic shortcuts in interpretation and choice. In the case of unconditional trust, "reassurance" and the "leap of faith" take place on the level of mode selection (!), the parameters of which display the individual's history of learning and socialization. Only in the case of conditional trust will

trustors consciously access the expectation of trustworthiness. In this case, the context determines the relevance of trust-related knowledge and enables, via appropriateness beliefs, the formation and generation of expectations of trustworthiness. The model of trust thus put forward incorporates and reductively explains both types of conditional and unconditional trust.

4.7. Theoretical and Empirical Implications

The model of trust and adaptive rationality bears a number of theoretical and empirical implications. In this section, I will derive a number of general propositions and prepare for an empirical test. As it is, the mode-selection threshold for the automatic selection of an action k, given frame i and script j, can be written as: $AW(A_k) = m_i * a_{j|i} * a_j * a_{k|j} > 1 - C / (p * (U_{rc}+C_w))$. Principally, to see how the mode-selection threshold is affected by a parameter change, one can vary the desired parameter, holding all other variables constant, and analyze its effect on the threshold. For example, a decrease in opportunity p will always lower the value of the right-hand side of the threshold. For a given match, mental effort, opportunity cost and motivation, this implies that the likelihood of a selection in the automatic mode increases. We will systematically develop such propositions in the following.[20]

To demonstrate how the model can be used to derive these general propositions, it is helpful to simplify the threshold using assumptions A1 and A2 (see chapter 4.6). Assume that we are looking at a social norm as a source of trust-related knowledge. If a norm is appropriate in the social situation which constitutes the trust problem, then the above threshold can be simplified, because (A1) the adopted trust frame F_i will unequivocally prompt to the script of the norm, so that the temporary accessibility of the normative script $a_{j|i}=1$, and (A2) the norm, will unequivocally regulate the course of action. For example, in the case of a reciprocity norm, a trustworthy response is normatively demanded, prompting to a trustful choice, and thus $a_{k|j}=1$. In this case, the choice of a trusting act is unconditional and rule-based whenever

$$m_i * a_j > 1 - C / (p * (U_{rc}+C_w))$$

Note assumptions A1 and A2 were already established when defining an ideal-type trust frame, which (1) unambiguously defines a trustee as trustworthy, (2) includes favorable expectation of trustworthiness and (3) prompts to the choice of a trusting act as a unique course of action. The trust frame F_i, by definition, involves $a_{j|i} = a_{k|j} =1$. From this model, a number of propositions can be derived. First, in analyzing the effect of the match m_i, it is easy to see that,

[20] Apart from propositions about simple main effects, the multiplicative link between the model parameters suggests that there exist a number of interaction effects. These will be developed in full detail in chapter 6.

Proposition 1 (ambiguity): The probability of an unconditional choice of a trusting act in a trust problem increases with the match m_i of a relevant trust frame i.

This proposition addresses the process of pattern-recognition in the stage of interpretation, and the aspect of "situational normality" as a basis for trust. If the situation is ambiguous and initial categorization yields only a low activation weight of the trust frame (a low match m_i), then the process of smooth pattern-recognition will be disturbed and actors will have to engage in a more elaborate process of interpretation. This also reduces the likelihood of a subsequent unconditional choice of a trusting act, since the match is carried over into the stage of action selection. Broadly speaking, proposition 1 establishes situational normality as a basis for trust, as mirrored in the question of a high *versus* low match of a trust frame.

Importantly, the subjective (un-)ambiguity represented in the match can have several reasons: actors may lack the appropriate knowledge to interpret the situation, they may be unable to link the available situational cues to the frames stored in memory, "noise" may make it difficult to interpret the cues, or relevant cues may themselves be absent or ambiguous. Thus, to be more precise, we can to decompose the match into its constituents. As it is, $m_i = a_i * l_i * o_i$ and a high match depends both on the chronic accessibility of a frame, its link to situational objects, and their unambiguous presence as an object in the situation. Thus, we can furthermore state that,

Proposition 1.1 (frame accessibility): The probability of an unconditional choice of a trusting act in a trust problem increases with the chronic accessibility a_i of a relevant trust frame.

Proposition 1.2 (cues): The probability of an unconditional choice of a trusting act in a trust problem increases with the presence of cues o_i indicating the appropriateness of relevant a trust frame.

Proposition 1.3 (link): The probability of an unconditional choice of a trusting act in a trust problem increases with the link l_i between salient situational objects and a relevant trust frame.

"Salience" here refers to salience in Higgins´ sense (see section 4.2.1). As we know, there can be internal (top-down) and external (bottom-up) reasons for a shift of attention to a particular stimulus. How precisely a situational object reaches the focus of attention and attains situational salience is, of course, an empirical question. The important point is that *some* objects will be perceived and thus be salient in the stage of interpretation, and *some* initial categorization will be made based on the spreading activation that these stimuli trigger. The link l_i refers to the "associative strength" (Fazio 2001, 2007) between situational objects and the mental model of the initial categorization.

Next, the model also predicts an effect of script-internalization. The parameter a_j captures the chronic accessibility of a relevant script; for example, that of a trust-related norm. If it is high, the script including the norm will be readily accessible given a definition of the situation that indicates its appropriateness. Thus,

Proposition 2 (script internalization): The probability of an unconditional choice of a trusting act in a trust problem increases with the chronic accessibility a_j of a relevant script.

In fact, since we have $a_{j|i} = a_{k|j} = 1$, the degree of internalization is a crucially decisive factor governing the conditionality or un-conditionality of trust. Of course, in a more general example, we would also have to address the question of temporary accessibility $a_{j|i}$ and, if the script was not a norm, its degree of regulation $a_{j|k}$. These have been ruled out in the simplifying example of an ideal-type trust frame; the corresponding additional propositions can be easily deduced.

Together, the match m_i of the frame and chronic accessibility a_j of the script define the left-hand side (the activation weight AW) of the mode-selection threshold. Further propositions can be derived when looking at the right-hand side and the remaining parameters. To begin with,

Proposition 3 (opportunity): The probability of an unconditional choice of a trusting act in a trust problem increases with decreasing opportunity p to activate the rational mode.

Most importantly, this addresses an aspect of situational opportunity in the form of time pressure and cognitive load and its related individual-intrinsic counterparts of cognitive capacity. For example, opportunity is low with high time-pressure or when there is high cognitive load (a concurrent processing of several tasks, for example). Similarly, opportunity can be low if thematic opportunity is absent, that is, if thematic knowledge with respect to a given decision problem is missing and the individual ability to make an appropriate judgment in a certain thematic domain is deemed insufficient. A lack of opportunity increases the likelihood of the activation of the automatic mode and use of heuristics. On the other hand, if opportunities do exist, then actors are more likely to engage in the rational mode because it is feasible. This proposition refers to both the stages of interpretation and choice alike.

Furthermore, the mental effort and costs C incurred with the activation of the rational mode are relevant to trust. From the threshold, we see that:

Proposition 4 (effort): The probability of an unconditional choice of a trusting act in a trust problem increases with the effort C associated with the activation of the rational mode.

Activating the rational mode always incurs some costs in the form of time and energy consumption. A situational factor that crucially influences this parameter is task complexity (Payne et al. 1992). A very complex task involves a large number of mental operations which have to be carried out in order to solve the problem at hand, increasing the perceived effort and mental costs associated with an elaborate processing mode. Note that task complexity can vary between different instances of a trust problem. Even when the necessary basic steps of interpretation do not add "excessive" demands on the cognitive system, the structure and complexity of the trust problem vary, for example with social embeddedness. Nevertheless, a rational elaboration during interpretation and choice will unescapably incur *some* effort which only a selection in the automatic mode can liberate the actors of.

There is also an individual-intrinsic aspect of effort and mental costs. It can be interpreted as being determined by individual cognitive ability. Different individuals may experience a different effort associated with the same task, simply because their cognitive abilities differ. Consequentially, they will associate different costs with a reflective reasoning process in the rational mode, both with respect to interpretation and choice. Low cognitive ability individuals may therefore be more prone to use unconditional trusting strategies. Thus, the cost parameter can also be interpreted as capturing an inter-individually stable difference:

Proposition 5 (motivation): The probability of an unconditional choice of a trusting act in a trust problem decreases with the motivation to activate the rational mode.

In the model, motivation is captured by the two components U_{rc} and C_f, which represent, respectively, the additional utility that the identification of a correct frame or action yields, given that the initial categorization is wrong (U_{rc}), and the disutility of making an inference error by following a wrong initial categorization (C_f). Jointly, they represent the opportunity cost of making an error and therefore translate into the motivation to engage in a more elaborate reasoning process. To abbreviate, simply write (U_{rc}+C_f) = U. The values of the parameters depend on the incentive structure of the trust problem and on the initial categorization that a trustor adopts.

An interesting implication of the model is the following: depending on whether the initial categorization suggests trust or distrust, that is, harbors a favorable or unfavorable expectation of trustworthiness, the trustor will attend to a different part of the incentive structure of the trust problem as a motivational basis during mode-selection. In short, a "trustful" and favorable initial categorization will push a trustor into focusing on the potential harm that a failure of unconditional trust can have, and the potential utility increase of withholding trust. On the other hand, a "distrustful" and unfavorable initial categorization will focus the trustor´s motivation around the potential benefits of a trustworthy response and the potential utility increase it affords as compared to the *status quo*. That is, different parts of the incentive structure will

become relevant to mode-selection in terms of motivation, depending on the initial assessment of trustworthiness (we can furthermore assume that the potential harms of failed trust always come to the attention of the trustor whenever an elaborated reasoning process has been initiated).

To see why this is the case, remember that the trustor can receive "reward" payoffs R if the trustee is trustworthy, incur "sucker" payoffs S if the trustee fails trust, and can opt for "punishment" outcomes P if he distrusts and maintains the *status quo* (the payoff relation is S<P<R).

It is easy to see that, if the initial categorization m_i defines the trustee as trustworthy (it provides a favorable assessment of trustworthiness), then $C_f = |P–S|$, which represents the experienced disutility relative to the *status quo* that the unconditional choice of a trusting act and a subsequent betrayal of trust yield to the trustor. Furthermore, $U_i = R$, that is, following the initial categorization yields the reward payoffs when the trustee is in fact trustworthy; and lastly, $U_{rc} = P$, because the trustor can reach the *status quo* payoffs and prevent the sucker payoff if he switches into an unfavorable assessment if the trustee is indeed not trustworthy. Thus, $U = (U_{rc} + C_f) = (P + |P\text{-}S|)$. Note that, for a favorable initial categorization, it is especially the potential harm S of a failure of trust and the *status quo* scenario P that matter as a motivational determinant of mode-selection.

In contrast, if the initial categorization suggests that the trustee *is not* trustworthy, then $C_f = |P–R|$, that is, the trustor makes a wrong decision by sticking to his initial categorization whenever the trustee is in fact trustworthy. The trustor then incurs an opportunity cost because he forfeits the potential gain R that he could have attained if he had not followed his initial categorization realizing the *status quo* payoffs P. Secondly, $U_i = P$, because the trustor can reach the *status quo* payoffs whenever he follows his initial categorization in a state of the world where the trustee is in fact not trustworthy. Lastly, $U_{rc} = R$, that is, the trustor can improve his utility from P to R if he revises his initial judgment and switches to a favorable assessment when the trustee is in fact trustworthy. All in all, we have $U = (U_{rc} + C_f) = (R + |P–R|)$. In other words, it is especially the potential gain R and the *status quo* scenario P that are relevant as a motivational basis for a trustor in the case of an unfavorable initial categorization.

R and S represent a situational aspect of motivation and what is "at stake" for the trustor. If the content of the trust relation is about an issue that is of high importance and promises high utility for the trustor, much can be gained, but much can also be lost. A failure of trust structurally involves "sucker" payoffs S which relatively put the trustor in a worse position than distrust and the maintenance of the *status quo* P (see chapter 3.3.3). Therefore, when the trust problem structurally involves a high utility-increase |P–R|, in the case of a trustworthy re-

sponse, and a high utility decrease |P–S|, in the case of failure, the activation of the rational mode and a subsequent conditional choice of a trusting act become more likely because the motivation to engage in the rational mode is high. Although we cannot know which initial categorization a trustor will adopt *a priori*, we can summarize both aspects of the incentive structure – the rewards R and the sucker payoffs S, relative to the *status quo*, as two important structural-situational components of cognitive motivation U.

There is more to this – relatively technical – argument. As we have seen, many authors argue that trust is often warranted as a "default" strategy (Luhmann 1979: 73, Jones & George 1998, Hill & O´Hara 2006, Keren 2007, Schul et al. 2008) because it is cognitively less costly to maintain, often culturally pre-defined as a socially shared rule and therefore preferred to initial distrust. At the same time, social psychological approaches sketch the human cognitive system as being built around a type of "default-interventionist" architecture (Kahnemann 2003, Evans 2008). Therefore, it may come to no surprise that most conceptualizations of trust emphasize the aspect of vulnerability over that of the potential gain involved in trust. If the above holds true, then our cognitive system is biased towards a potential detection of the harms involved in the trust problem, because a trustworthy initial categorization is adopted "by default". Structurally, it is the potential loss we can incur that is relevant to the mode-selection of interpretation and choice, given that we start from a default assumption of trustworthiness. A number of experimental studies have been concerned with the effects of "stake sizes" on decision-making and problem solving, and the results overall point into a direction that conforms to proposition P5.[21]

Propositions P1-P5 concentrate on main-effects that can be derived from a comparative-static analysis of the model without looking at any interactive effects between the variables. In addition, a number of corollary propositions can be derived from P1-P5. To begin with, a trust-frame, in the ideal-type case, entails a favorable expectation of trustworthiness. If the frame is unambiguously valid, then there is no doubt about the trustworthiness of the trustee. In other words, the degree of ambiguity experienced in the situation by the trustor will directly influence the expectation of trustworthiness, and the match m_i can be regarded as a direct equivalent of the expectation of trustworthiness. If a trust frame is unambiguously valid, then expectations of trustworthiness can be stabilized at a favorable value of the expectation.

Proposition 6 (interpretation/expectation): The trustor´s expectation of trustworthiness increases with the match m_i of a relevant trust frame.

[21] This issue is more fully explored in section 6.2.3 below.

The pattern-recognition of stored mental models and the natural assessment of the match and processing fluency involve the context-sensitive application of knowledge under the headnote of adaptive rationality. Thus, P6 reformulates the suggestion that frames could explain the formation of initial beliefs (in that "frames move beliefs move choice", Batigalli & Dufwenberg 2009) and it delivers an explanation for the context-dependency of these initial beliefs. However, this observation is traced back to the "appropriateness" of trust-related knowledge and use of endogenous adaptive rationality. In addition to assuming that choice is preceded by a framing stage, the model states that framing simultaneously influences the subsequent degree of rationality. At the same time, the match defines the appropriateness beliefs of the trustee, and therefore fixes expectations and second-order beliefs. This "analogy" between the match and expectations can be established because we have *defined* a relevant trust-frame to include a favorable expectation of trustworthiness. In this way, appropriateness and expectations coincide. If the trustor switches to the rational mode, then the expectation of trustworthiness will be a consciously perceived representation of the appropriateness belief p_i.

We also have seen that unconditional selections (of frames, scripts, actions) are based on the re-cognition of stored mental schemata and their routine automatic application, in which a direct link from associative memory to behavior, *via* spreading activation and respective activation weights, does exist. An automatic selection of an action describes a completely different "logic of selection" than that of an instrumental choice. The automatic use of stored schemata follows the activation weights, whereby the influence of instrumental variables is suppressed. This can statistically be interpreted as a negative interaction effect between the parameters of mode-selection and instrumental (rational-choice related) variables. Thus,

Proposition 7 (instrumental variables): The influence of instrumental variables (first- and second-order beliefs, incentives etc.) in a trust problem decreases with the match m_i of a relevant trust frame.

Lastly, we can address the issue of discriminating between automatic and rational selections in general. On the surface, conditional and unconditional trust result in the same outcome; in both cases, the observable overt behavior is the choice of a trusting act and a transfer of control or resources to the trustee. However, we can hypothesize that the use of mental "shortcuts" does in fact have an effect on the time that a trustor needs to reflect upon the trust problem in order to make a decision about the choice of a trusting act. An elaborated reasoning process inescapably uses up *some* time, and should (on average, even if the differences are minimal), take longer than a blind and unconditional choice based purely on automatic processing. This hypothesis is also supported by empirical studies showing that unconditional trusting strategies have a lower decision time than conditional trusting strategies (Krueger et al. 2007). In short, decisions based on automatic information processing should be faster as compared to deliberative decisions because heuristics are quickly accessible, and rational

cognitive operations are time-consuming. As Schunk and Betsch put it, "maximizing is a highly cognitive process, involving conscious weighting, and information search, for example, which requires more cognitive capacity than affective-intuitive, satisfying decisions" (Schunk & Betsch 2006: 394).

Proposition 8.1 (decision time): The decision times using the automatic mode are shorter than the decision times using the rational mode.

Proposition 8.2 (trusting strategies): Unconditional trust has a shorter decision time than conditional trust.

Propositions P1-P8 can be translated into a set of empirical hypotheses and tested using an experimental design. If the model is applied to other problems than that of an ideal-type trust-frame, the specification of the model parameters (i.e. our simplification $a_{j|I} = a_{k|j} = 1$) may need to be adjusted, and additional propositions can be added for these parameters accordingly.

So far, we have only looked at propositions of main effects. A number of additional and very interesting model propositions can be generated when more than one parameter of the threshold is varied simultaneously. As the model suggests, the final "balance" of the left- and right-hand side of the mode-selection threshold jointly depends on the value of all parameters involved. This means that the effect of a change in one variable depends on the value of another. For example, it is easy to see that the effect of cognitive motivation can be compensated by a high match. Statistically, this translates into a predicted interaction between motivation and chronic accessibility, or any other component of the match, for that matter. A high match may completely suppress and counter-balance the effect of cognitive motivation. Likewise, we have to expect an interaction between motivation and opportunity, between motivation and effort, as well as between effort and opportunity, and so forth. As it turns out, a number of specific interaction effects are a model-inherent feature that harbors a set of predictions which contrast both standard psychological and economic models. These interaction hypotheses will be fully developed in chapter 6.3 below. At this point, we formulate the following general proposition:

Proposition 9 (parameter interactions): All model parameters simultaneously define the mode-selection threshold value. The effect of a change in one parameter depends on the value of all other parameters that simultaneously define the threshold. Therefore, all parameters of the mode-selection threshold are connected in second- and higher-order interactions.

Overall, the model of trust and adaptive rationality developed so far helps to spell out the conditions of conditional and unconditional trust. It directs our attention towards the parameters of mode-selection, which have to be understood as the primary cause for a "leap of faith" in trust. As propositions P1-P9 demonstrate, a number of important theoretical stipulations can

be derived from a static-comparative analysis of the mode-selection threshold. Chapter 6 of the book presents an attempt of an empirical experimental corroboration, including a specification of testable hypotheses which make use of propositions P1-P9.

5. The Social Construction of Trust

"More or less consciously, agents can contribute to the development of the trust-inducing contexts, which, in turn, enable them to trust more easily" (Möllering 2005: 6).

In the preceding chapters, we have approached the phenomenon of trust from the trustor's perspective, relating it to individual adaptive rationality, interpretation, and choice. In other words, we have restricted ourselves to an analysis of trust as residing in the psychological state of the trustor who "passively" responds to the environment and seeks a solution to the trust problem. Importantly, the immediate situation and its context define the relevance of trust-related knowledge. By providing the situational objects and cues that govern the processing mode and trigger the activation of associated frames and scripts, they serve as a basis for interpretation and choice and the context dependent adjustment of rationality.

In this chapter, we will take a look at how the trust relation, as a social system, is "actively" constituted by the actors involved. Far from being a passive achievement, interpretation and the subjective definition of the situation are normally reached in symbolic interaction with others, and rely on a dynamic process of communication. At the same time, communication is at the root of the constitution of social systems. Any social system can be reconstructed as a genetic sequence of meaningful communications, in which the actors' subjective definitions of the situation temporarily converge into a shared *social* definition of the situation. This process of *social framing* reflexively structures the situation, and also the context. Actors use a socially shared stock of knowledge to interpret situations, and in doing so, they externalize meaningful symbols that confirm its appropriateness. Therefore, social structure continuously reproduces itself in a reflexive process of structuration and "agency." Concerning trust, this suggests that (1) trustor and trustee actively constitute the social system of a trust relation in a process of social framing, and (2) the context of the trust relation is not static, but highly dynamic and shaped by the actors involved.

Furthermore, the media which actors use to communicate (language, writing, generalized media of exchange, etc.) differ with respect to their abilities to transmit meaning and their capabilities to reduce ambiguity or overcome situational constraints. Many trust researchers argue that face-to-face communication is the most effective means of socially framing a trust problem, because it provides a very rich set of cues which trustors can make use of. On the contrary, "lean" media restrict information and are not conducive to a build-up of trust, because they convey fewer cues, increase anonymity, and open up the potential of defection. Apart from verbal and nonverbal communication, the tangible actions involved—the choice of a trusting act and its (un)trustworthy response—can become a significant symbol for facilitating social framing as well. All in all, a number of different signals convey how the parties "view" the status of the relationship.

The analysis of communication media points to another aspect: communication is often "relational" and addresses the relationship orientations of the actors involved. Such relational communication defines and changes the status of a relationship. This information is coded in the relational schemata which actors use to define the trust problem. As will be argued, relational schemata constitute a very important class of trust-related knowledge, because most of our social relations are in fact dyadic, and we often use relational schemata in "transference," even when a particular significant other is absent. In addition, relational schemata are hierarchically structured and exist even for highly generalized types, such as interactions with "a stranger." Furthermore, not only do relational schemata include information about the status of the relationship and of the interaction partner, but they also define the identity of the actor within that relationship. This means that interpretation does not only concern the definition of "external" elements, but also concerns the actor's self-concept, or identity. In short, during social framing, actors reciprocally define both their identity and social identity as well.

The concept of identity can be fruitfully connected to trust research. Firstly, personal identity comprises the stable "traits" and disposition to trust, which we have discussed in chapter 3.1. Secondly, the actor's social identity comprises collective and relational self-concepts, which also serve as a springboard for trust and motivate the choice of a trusting act. For instance, trust researchers have argued that trust can be based on a salient group identity and collective identification with other social aggregates. In line with several identity theory approaches, these models hold that a salient social identity triggers a shift in the level of self-identification, leading to a focus on aggregate-level goals. Likewise, "in-group favoritism" and categorical attributions of trustworthiness in the form of "stereotyping" can aid and support the choice of a trusting act. Relational identities which are included in, and framed by, relational schemata, achieve similar effects with respect to dyadic relations with significant others. An important mechanism behind these effects is "self-verification" and the confirmation of an adopted identity, whereas a mismatch between the standards of identity adopted and those confirmed by others often triggers defensive behaviors and distrust.

This naturally leads to the question of which identities actors adopt in a situation. Since identities are activated during the subjective definition of the situation, they are equally subject to symbolic interaction and communication events. At the same time, this means that actors can intentionally try to mimic a false identity—for example, one that is known to be associated with a trustworthy reputation. Thus, the process of "identity signaling" constitutes an important stage that precedes the conventional trust problem. All in all, the perspective we develop emphasizes that the social construction of trust is actively achieved in a process of communication between trustor and trustee, in which conditions favorable for a build-up of trust are produced.

In the last sections of this chapter, we will take a look at "active trust," and the strategies which trustor and trustee may use to manage trust and trustworthiness. Active trust implies that the performances of the actors involved, and the actions they take to produce trust, enter the focus of interest. One paradigm that can inform such a perspective is "impression management" research, which is unequivocally concerned with how individuals can manage their self-presentation and the impressions they convey to others. Trust, in this sense, can be actively produced by different "performative strategies" of self-presentation, such as increasing the other's commitment, showing similarity, displaying trustworthy characteristics, managing the other's emotional threats, or producing the appearance of situational normality. All in all, the discussion of active trust development, impression management, and the particular trust management strategies which can be applied completes the picture of the social framing perspective. It adds to the "logic of explanation" of the trust phenomenon the last step of a micro-macro transition to the collective outcome of a trust relation. The discussion shows that trust has to be understood as an ongoing process of reflexive structuration in which both trustor and trustee—sometimes intentionally, sometimes automatically—achieve a shared definition of the situation, favorable expectations, and a confident choice of a trusting act.

5.1. Defining the Context

5.1.1. Symbolic Interaction

The model of adaptive trust developed in the last chapter combines objective and subjective elements of the situation in the process of framing, in order to explain conditional and unconditional types of interpersonal trust. The processes of mode selection and interpretation depend on the availability and accessibility of mental schemata, as well as on the presence of situational cues which serve as indicators of the appropriateness of a particular trust frame. The match between frames and cues available in the environment decisively influences the activation of trust-related schemata, the selection of the processing mode, and the particular definition of the situation an actor adopts.

Importantly, frames do contain associative knowledge about the situational *cues* which serve as a trigger for their activation. In an environment that is primarily social, it is reasonable to assume that a majority of these cues consist of actions of other actors. Their overt behavior is not only an observable objective "fact," but also represents a subjectively meaningful *symbol* with a cognitive, expressive, and appellative function to an observer. If an actor can decipher the meaning of an action—that is, recognize it as an indicator of the other's intentions and as

an expression of the other's situational definition—it is a *significant symbol* (Mead 1967).[1] Importantly, a significant symbol objectifies the perspective of its sender within the social environment and conveys his idea of a future course of action, his intentions, and his meaning to those who can observe it and interpret it correctly. In this way, significant symbols form a basis on which other actors can adjust their own definition of the situation.

If actors, intentionally or unintentionally, influence each other's situational definitions by expressing purpose and meaning through significant symbols to reach a social definition of the situation, we speak of *symbolic interaction* (Blumer 1969, 1974). During symbolic interaction, actors utilize a number of different *media* (or "symbol systems," such as language, writing, and generalized media of exchange), which systematically convey a specific, culturally and institutionally determined meaning. Media therefore provide a vast repertoire of ready-made significant symbols that can be used to indicate one's own perspective to others. Because they effectively transfer meaning, media have the power to influence others' interpretations and define their "appropriate" frames of reference in a particular context. Ultimately, a shared understanding of meaning and the social definition of the situation, reached during symbolic interaction lay the ground for cooperation. In the words of Blumer, "the fitting together of lines of conduct is done through the dual process of definition and interpretation ... established patterns of group life exist and persist only through the continued use of the same schemes of interpretation; and such schemes of interpretation are maintained only through their continued confirmation by the defining acts of others" (Blumer 1966: 538). In doing so, individuals symbolically influence each other's interpretations until they interlock in a congruence of attributed meaning. Symbolic interaction allows for a flexible coordination of action because the actors, by making use of media, adjust their interpretations empathically and converge on a shared definition of the situation.

However, a social definition of the situation does not occur immediately and by itself. It is actively produced by the actors in a reflexive process of communication. Quite generally, communication can be regarded as "the mechanism through which human relations exist and develop—all the *symbols* of the mind, *together with the means of conveying them* through space and preserving them in time. It includes the expression of the face, attitude and gesture, the tones of the voice, words, writing, printing, railways, telegraphs, and whatever else may be the latest achievement in the conquest of space and time" (Cooley 1983: 61, emphasis added). In other words, *communication* can be defined as a shorthand for symbolic interaction with the help of media (Esser 2000a: 248). Communication is always selective: a sender has

[1] Mead more generally regarded any *gesture*—"those phases of the [social] act which bring about adjustment of the response of the other" (Mead 1967: 44) as significant symbol, if it conveyed an *idea* about a future course of action: "When, now that gesture means this idea behind it [shaking a fist to indicate the idea of a possible attack] and it arouses the idea in the other individual, we have a significant symbol" (ibid.).

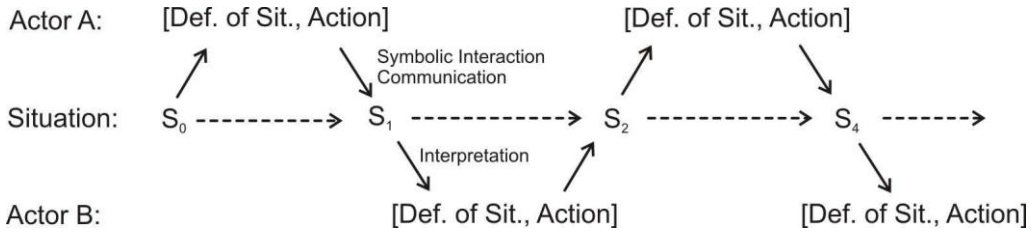
to select information from a repertoire of "open possibilities" (Luhmann 1995: 140), formulate a message, and choose the medium by means of which it will be transmitted. The receiver has then to receive, decode, and understand its meaning. Communications may fail when a wrong selection is made by the sender, by the receiver, or by both. Normally, actors are aware of the openness and selectivity inherent in the process of communication. Taken together, the successful emission, transmission, and reception of a message (including the correct decoding of its meaning) constitute an *elementary unit of communication*. For communication to continue, several elementary units must sequentially connect to each other. In other words, a meaningful continuation of communication requires some opportunity to follow up with new elementary units of communication and a successful linkage to past sequences.

To denote the fact that a link between elementary units of communication may be successfully achieved, Luhmann (1990, 1995: 137ff.) uses the term *structural coupling*. Importantly, he argues that social systems must be understood as a continuous aligning sequence of elementary units of communication. The constitution of social systems rests on a sequential structural coupling of elementary units of communication, and requires a "synthesis of information, utterance and understanding (including misunderstanding)" which "has to be recreated from situation to situation by referring to previous communications and to possibilities of further communications which are not restricted by the actual event" (1990: 3). Elementary units of communication must convey the minimal meaning necessary for reference by further communication and for a continuation of the social system. Frames and scripts—the cultural and normative stock of knowledge shared by the actors—facilitate this continuation by providing answers to the questions, "What kind of situation is this?" and, "What am I supposed to do?" Actors can thus easily decode the meaning of the typical actions and typical situations they encounter; they routinely interpret the typical significant symbols communicated. Stable social systems are temporarily constituted on this foundation of routine cultural knowledge which enables the "structuration" (Giddens 1984) of social systems and the "agency" (Emirbayer & Mische 1998) of human behavior.

Any social system can be reconstructed as a genetic sequence of meaningful communication, which the actors, as the "personal systems," initiate and sustain using the "cultural system" of internalized frames and scripts to aid them. Within the MFS framework, we can understand the constitution of social systems as a process of *social framing* (Esser 2001: 496). Social framing describes sequences of individual frame selection and action selection, their aggregation into a new objective social situation, and feedback into new individual framing processes. Actors interpret situations and "externalize," or communicate, their intentions and interpretations in the form of actions and significant symbols. This changes the objective situation for other actors involved, and feeds back into the next sequence of individual framing and choice. In this way, social systems are endogenously and "reflexively" created by the actors; and

communication guides individual framing into a temporary convergence of attributed meaning (figure 18):

Figure 18: Communication and social framing



A trust relation is clearly a social system. Its emergence and continuation also depends on structural coupling and the temporary convergence of communicated meaning. If achieved by the actors involved, this is expressed in the (confident) choice of a trusting act and a trustworthy response. The trust relation as a social system is "locally" constituted within a particular social environment as result of social framing sequences. It is guided by the application of shared interpretive schemata, which are reflexively activated during communication. Unsurprisingly, a number of trust researchers have argued that trust must be traced back to a symbolically negotiated social definition of the situation, to an (implicit or explicit) orientation of trustor and trustee which rests on a shared (implicit or explicit) understanding of the logic of the situation, and which constitutes and symbolizes itself in the "meaningful" choice of a trusting act and its trustworthy response (Jones & George 1998, Endress 2002, Kramer 2006, Möllering 2006).

At the same time, the constitution of a trust relation is tightly connected to the constraints and properties of individual adaptive rationality during the framing processes. The attribution of meaning and the meaningful continuation of communicative sequences both rely on a constant interpretive effort on the part of the actors involved. For unconditional trust to emerge, the chains of communication associated with a trustful course of action need to unfold without problematic interruptions, and significant symbols must be effortlessly decoded, so that a structural coupling of communicative acts smoothly accumulate into the choice of a trusting act and its trustworthy response.

So far we have approached the phenomenon of trust from the trustor's perspective, relating it to *individual* framing, adaptive rationality, and information processing. In other words, we have restricted ourselves to an analysis of trust as residing in the psychological state of the trustor who passively responds to the environment. But the above arguments demonstrate that communication is decisive for the development of trust, because it actively defines the context in which those individual framing processes occur. Möllering uses the term "reflexive structuration" (2006: 99) to describe this endogenous feedback between the context, communica-

tions, interpretation, and choice in all trust problems. A symbolic-interactionist perspective on interpersonal trust suggests that trust relations must always be reciprocally and actively defined, in that communication serves as the springboard for interpretation. In short, the context of the trust relation is not static. It is highly dynamic and endogenously shaped by the actors involved, by their actions and communication.

5.1.2. Language and other Signals

The channels on which elementary units of communication are transmitted differ with respect to their ability to overcome situational constraints (such as time, location, distance, and permanence), resolve ambiguity, and convey symbolical meaning. Consequently, they are often classified according to their *richness* (Daft & Lengel 1984, 1986), which describes the "varying capacities for resolving ambiguity, meeting interpretation needs, and transmitting data" (Trevino et al. 1987: 557).[2] For example, direct face-to-face interaction is considered to be a particularly "rich" channel, because it allows for instant audio-visual feedback and delivers a multitude of symbolical cues (vocal tone, emotional expression) to the actors. That is, dyadic face-to-face interactions possess a particularly high amount of symbolic content (Daft & Lengel 1986). On the other hand, "lean" channels are more indirect (a written email, a television broadcast, a phone call), generally emit fewer cues, restrict feedback, and are less effective in resolving ambiguity and equivocality (ibid.). Overall, communication channels "differ in the extent to which they are able to bridge different frames of reference, make issues less ambiguous, or provide opportunities for learning in a given time interval" (Rice 1992: 477).

With respect to interpersonal trust, this suggests that channel richness is an influential factor determining the build-up of trust (Hollingshead 1996, Meyerson et al. 1996, Alge et al. 2003). For example, Rockmann and Northcraft (2008) propose that the richness of the communication channel directly affects the build-up of both cognition-based and affect-based trust, as it is related to the frequency of deception and defection. For one, lean media encourage defection because "they offer fewer social context cues, driving individuals to feel more anonymous" (ibid. 108), and also because the deceiver has to control fewer potential "leakages." On the other hand, "as the richness of the medium increases, the potential deceiver would need to control more aspects of his or her communication to be successful in the deception attempt" (ibid.). In sum, they state that rich media, by providing multiple cues and information, facilitate conditions conducive to the build-up of trust, whereas lean media make it more difficult to gather information, and encourage opportunistic behavior. Empirically, they find that face-to-face communication generates the highest levels of trust and trustworthiness, as compared

[2] Communication researchers commonly use the term "media richness" to refer to the technical and informational constraints pertaining to different communication channels. The term "media" is used by then in a slightly different meaning from that adopted here, i.e. it does not refer to different symbol systems, but to the different technical channels which can carry the symbol systems.

to communication mediated by video or computer (see also Valley et al. 1998, Jarvenpaa & Leidner 1999, Alge et al. 2003, Wilson et al. 2006).

These results are in line with a large body of empirical work that has examined the effect of communication in a variety of social dilemmas. Quite generally speaking, communication via language has been found to be a prime factor boosting trust development, and the experimental results showing that it enables trust and cooperation have, over the decades, accumulated into an enormous bulk of evidence (see, for example, Loomis 1959, Dawes et al. 1977, Isaac & Walker 1988, Orbell et al. 1988, Sally 1995, Ostrom 2000, Bicchieri 2002, Malhotra & Murnighan 2002, Ostrom 2003, Charness & Dufwenberg 2006). Notably, face-to-face interactions are usually found to be much more effective than any other form of interaction in inducing trust. Nevertheless, communication may indeed become "cheap talk" (Farrell 1987) when the messages and the media channels become so sparse that only minimal information can be transmitted (Bracht & Feltovich 2009). Normally, however, human language offers an endless repertoire of significant symbols to convey one's perspective to others. It is the most important medium of communication in the development of trust and for the successful constitution of a trust relation. The interpretation of linguistic symbols is therefore a fundamental aspect of trust development and a primary means of overcoming trust problems (Bacharach & Gambetta 2001).

A particularly effective variant comes in the form of promises. Promises refer to "obliging" yet nonbinding commitments of the trustee to reciprocate the risky investment of the trustor. At the same time, they represent a direct "invitation" to choose a trusting act. In contrast to the "cheap talk" predictions of rational choice, promises are often very successful in convincing the trustor of one's trustworthiness and initiating the choice of a trusting act. Presumably, this is because the normative obligations that reverberate in trust and trustworthy response are made explicit in the promise. Empirically, promises have been found to induce trust even in "one-shot" situations (Charness & Dufwenberg 2006). However, as Charness and Dufwenberg note, in making a promise, it is critical that trustees communicate personal intentions and formulate their promises in a personal style.

Many scholars contend that *nonverbal communication* (or "body language", see DePaulo 1992, Burgoon & Hobbler 2002) is of equal importance.[3] For one, nonverbal communication

[3] The term *nonverbal communication* subsumes a variety of communication channels that serve to convey symbolic information other than explicit language, including *kinesics* (visual body movements such as gestures, facial expression, posture gaze), *paralanguage* (vocal cues other than the words themselves, such as pitch, loudness, tone), *physical appearance* (manipulable features such as clothing, hairstyle etc.), *haptics* (the use of touch, frequency, intensity and type of contact), *proxemics* (the use of interpersonal distance and spacing relationships), *chronemics* (use of time in messages, such as punctuality or waiting time) and *artefacts*, that is, manipulable objects of the environment that may convey messages from their designers or users (Burgoon & Hobbler 2002). Depending on the particular research method, communication re-

"qualifies" verbal communication by providing a number of additional cues which may reinforce and augment, or disprove and contradict the meaning communicated by language, and can have a "persuasive impact" on observers by influencing source credibility (Burgoon et al. 1990). For example, it has been found that prolonged eye-contact, a relaxed body posture, fluent speech, and a calm voice increase message credibility and the perceived trustworthiness of a sender (ibid.). They are therefore conducive to a build-up of favorable expectations of trustworthiness.

But nonverbal communication can also serve as a symbolic cue in its own right. The most important function in this regard is the expression of emotions and affective states, which is, to a substantial degree, "hard-wired" and automatically occurring (DePaulo 1992). Likewise, the recognition of faces and the processing of emotional displays is highly automatic (Todorov et al. 2009), constituting an important evolutionary adaptation for "threat detection" (Adolphs 2003). Generally speaking, displays of anger and sadness influence trustworthiness judgments negatively, while displays of happiness (i.e. smiling) increase perceived trustworthiness (Winston et al. 2002, Eckel & Wilson 2003). At the same time, the display of emotions is highly culturally regulated, and "each culture has deeply ingrained anticipations how, when, where and with what consequences emotions are displayed in public and private" (Burgoon 1993). Ekman (1972) used the notion of *display rules* to denote "cultural norms governing the management of emotional expressions [that] indicate which emotions should be conveyed, depending on the situation, the person who is communicating the emotion, and the person to whom the emotion is being communicated" (DePauolo 1992: 209). In other words, the frames and scripts stored in memory also contain information about emotional display and appropriate responses, and a nonverbal signal may well serve as a significant cue for the activation of trust-related schemata.

Other than that, many aspects of nonverbal communication can be, at least indirectly, related to judgments of trustworthiness as well. It is, for example, well established that a host of nonverbal cues (both static appearance and dynamic factors, such as expressivity, gaze, immediacy and involvement, and paralanguage) influence judgments of attractiveness (the "visual primacy effect" and the "what-is-beautiful-is-good" heuristic, see Eagly et al. 1991, Feingold 1992, Langlois et al. 2000). At the same time, judgments of attractiveness are directly related to judgments of trustworthiness (Wilson & Eckel 2006). Williams (2007) argues that trustor and trustee use a number of behavioral strategies to regulate self-presentation and perceived trustworthiness, claiming that nonverbal communication is a primary means for actively achieving "threat reduction." Similarly, Bacharach and Gambetta (2001) hold that nonverbal

searchers estimate the impact of nonverbal communication on the total "meaning" produced in communication to be between sixty and ninety percent (ibid.).

cues may be an important class of signals indicating "trust-warranting properties" because they are hard to imitate. Overall, nonverbal communication must be regarded as an important part of trust-related communication. When assessing the trust-related qualities of a trustee, nonverbal cues are often the most immediately available, and therefore can be used heuristically when more individuating information does not exist (Burgoon & Hobbler 2002).

But meaning is not only conveyed through language and nonverbal communication. The choice of a trusting act itself can be a significant symbol to communicate one's interpretation of a trust problem. In the economic models to trust development, this idea was expressed in the idea of "psychological forward induction"—the trustor uses his action to induce an update in the second-order beliefs of the trustee, who by the very fact of observing a trusting choice, can infer something about the beliefs of the trustor. Knowing this, the trustor can use his trusting act strategically as a signal to communicate his definition of the situation; for example, to induce "guilt" in the trustee and to secure a trustworthy response. In chapter 2.2.3, the very same idea was discussed in terms of the "moral obligations" that accompany the choice of a trusting act. Trust researchers have regularly expressed the power of the trusting act to define the situation and induce trustworthiness, even when rational grounds for favorable expectations do not exist. According to Gambetta, actors are able to learn "that it can be rewarding to behave as if we trusted even in unpromising situations" (1988a: 228), and Hardin similarly claims that "as-if trust can be willed repeatedly so that one may slowly develop optimistic trust" (1993: 515). The notion of conditional trust in which actors engage in only a "pretense" of suspension (Jones & George 1998) precisely points to this power of trust to create the behavior on which it ostensibly rests. In fact, most models of trust development implicitly include the assumption that trust starts from a very narrow basis in which actors, even "irrationally" and against their expectations, "just do it" (Möllering 2006: 115f.) and opt for the trusting act in order to test whether a trust relation is feasible.

A similar argument can be made with respect to the actions of the trustee, which are important signals to the trustor influencing the constitution (and reproduction) of the trust relation as a social system. For one, different forms of commitment and "hostage posting" can help to define the trust problem and influence perceived trustworthiness (see chapter 3.3.4 already). Principal-agent theory has proven a valuable tool in carving out the conditions that signals must meet in order to be reliable and to create a separating equilibrium in which there is no "mimicry," so that trustworthy trustees can use commitment to credibly signal trustworthiness (Raub 2004, Bracht & Feltovich 2008). As Raub puts it, "hostages that serve signaling purposes contribute to the 'definition of the situation' and to 'framing'" (2004: 344). In other words, the trustee invests resources into a signal that can credibly communicate his trustworthiness and benign intentions. In this way, his commitments become a significant symbol for

the trustor once he adjusts his interpretation of the trust problem accordingly, allowing him to confidently choose a trusting act.

Apart from that, the trustworthy response of the trustee (or the failure of trust) is a significant symbol in its own right, which strongly affects the future of the trust relation. When looking at developmental models in chapter 3.1.4, we discussed the idea that trust gradually evolves over time with successful ongoing social exchanges. Apart from "realizing" the content of the trust relation, a trustworthy response also confirms the appropriateness of the trust-related knowledge which was used to solve the trust problem. The trustor can infer that his definition of the situation was correct. For example, if a trustee was judged to be trustworthy based on an assessment of his benevolence and integrity, a trustworthy response will confirm the judgment of these characteristics. Likewise, if trust was afforded based on knowledge of the normative-institutional environment (say, a social role), a trustworthy response confirms the applicability and appropriateness of the script in the particular context. Most generally speaking, the trustee's response initiates learning and reinforcement of existing mental models, and this increases the associative strength between the perceived situational stimuli and the applied trust-related mental schemata.[4]

A failure of trust has a comparable symbolic consequence. Most trust researchers agree that a breach of trust is a disruptive event, substantially redefining the perspective of the trustor and impacting the future reproduction of a trust relation. What is more, researchers commonly claim that trust is "intrinsically fragile" (Gambetta 1988a) and is easier to destroy than to create (Barber 1983, Baier 1986, Slovic 1993, Meyerson et al. 1996, Robinson 1996). According to Slovic (1993), a variety of cognitive factors contribute to this asymmetry between trust-building and trust-destroying events. First, a failure of trust is more visible than a trustworthy response. This is because the violation of trustworthiness expectations triggers immediate arousal, an effect of the hot emotional charge that trust-related expectations possess. Second, failures of trust do have more weight in judgment than trustworthy responses. Slovic empirically demonstrates this asymmetry by presenting hypothetical (positive and negative) news to subjects. In support of the claim, he finds that negative news have more impact on judgments of trustworthiness than positive news. Similarly, Burt and Knez (1995) find that third-party information amplifies distrust to a greater extent than trust. Overall, these findings support the view that a breach of trust has a strong symbolic meaning to the trustor, and this breach negatively impacts on future trust.

[4] Within the theoretical framework of the MFS, the confirmation of the trustor's situational definition by a trustworthy response can be mirrored in a change in the (chronic and nonchronic) accessibility parameters of trust-related frames and scripts, in the links between both situational objects and frame, as well as in the associations between frames and scripts and scripts and actions ("encoding"). Presumably, if trust develops over time and actors gradually build up specific, trust-related knowledge structures, then all parameters are affected at the same time.

Lewicki and Bunker (1995a, 1996) focus on the question of the fragility of trust and develop a more sophisticated argument, holding that a violation of trust has different effects depending on the stage of trust development. Thus, the attributions that trustors will make about the failure of trust vary with the "basis" of trust—the fragility of trust decreases with relationship development, because with identification-based trust, violations may "easily repaired through the strong bonds that the parties have built with each other" (ibid. 168, see also Rempel et al. 1985, 2001).

More recently, scholars have focused on the conditions that facilitate or prevent trust repair (Dirks et al. 2009, Kramer & Lewicki 2010), the attributional processes by which a failure of trust obtains the symbolic meaning of a serious "transgression" (Tomlinson & Mayer 2009), and have started to spell out the precise types of violations and the interaction rituals necessary to restore them (Ren & Gray 2009). We will have a closer look at such interactional aspects further below—at this point, it is important to note that the trustworthy response itself is high in symbolic meaning, and will be evaluated by the trustor from exactly that interpretive perspective which was the starting ground for his choice of a trusting act.

Overall, it is apparent that the communication of symbolical content is central to the solution of trust problems. It is the power of communication to transfer meaning and enable mutual perspective-taking via significant symbols that makes it a most decisive part in the development of trust. A trust relation can be successfully established when trustor and trustee empathically converge in their situational definitions of the trust problem, and it can be maintained when observable actions match these interpretations. Communication is inseparably tied to this process of social framing.

5.1.3. Relational Communication

If social systems are sequences of communicative acts, then, ultimately, communication must be regarded as the "carrier" of trust per se. Only a congruence of perspectives and a convergence of meaning in both the trustor and trustee can lead to a successful constitution of a trust relation. Any communication can become a criterion for the continuation or termination of the trust relation, and every action can potentially create, sustain, or bring into doubt a corresponding trust frame. Luhmann consequently argues that,

> "In addition to its immediate significance as regards situation and purpose, every socially comprehensible action also involves the actor's presenting himself in terms of trustworthiness. Whether or not the actor has this implication in mind—whether he is aiming at it, or consciously disclaiming it—the question of trust hovers around every interaction, and the way in which the self is presented is the means by which decisions about it are attained" (Luhmann 1979: 39).

Humans are experienced as "a complex of symbols" (ibid.), continuously and often unconsciously expressing meaning, which others use as indicators for the judgment of trust-related characteristics and an assessment of trustworthiness expectations. What is more, communica-

tion creates and changes the attributions that actors hold toward each other and toward the relation between them. It does not only convey "referential" meaning, but also "relational" meaning, which enables individuals to interpret and define their relations (Watzlawick et al. 1967, Millar & Rogers 1976). In short, during interaction, actors not only reciprocally define "situations," but also define the status of their interpersonal relationships. Any aspect of communication that affects the current and future status (i.e. development, stabilization, or change) of a relation is *relational communication* (Millar & Rogers 1976, Burgoon & Hale 1984, 1987, Dillard et al. 1999). It is often implicit and embodied in the nonverbal signals that actors emit, but it can also be overt, deliberate, and intentional.

With relational communication, individuals impart to one another how they have defined the relationship and how they view themselves and the other within the relationship. In other words, they symbolize their *relational perceptions*, that is, those cognitions that refer to the status and quality of the interpersonal relationship. The choice of a trusting act and trustworthy response are examples of relational communications that directly affect the definition of the trust problem, but "mediate" verbal or nonverbal communication is of equal significance. Relational communication is critical to the development of interpersonal trust, because the majority of the cues which enable a trustor to make a "leap of faith" are contained within the relational signals emitted during interaction (Holmes 1991, Lindenberg 2000, 2003, Six 2005: 21f.).

Relational communication is effective on different "generic themes" which define the content of relational perceptions. These themes include, among others, dominance, affection, emotional arousal, formality, intimacy, involvement, composure (self-control), similarity, inclusion, and depth (Burgoon & Hale 1984, 1987). Generalizing these themes, Barry and Crant (2000) discuss four distinct dimensions on which relational perception occurs, and on which relational communication can consequently be effective: (1) dependence, that is, the extent to which dyad members depend on each other in relative comparison, (2) commitment, that is, the psychological attachment to the other and the intention to maintain the relationship, (3) transferability, that is, the existence of alternatives and "exit" options, mitigating the potential for exploitation, and (4) "confidence," that is, the perception that one will not be betrayed by the other in the future (ibid.).[5] In contrast, Dillard et al. (1999: 58) argue that the "basic substance" of all relational judgments, which also reflects "the fundamental phenomenological content of interpersonal relationships," can be reduced to two factors: (1) dominance, that is, the degree to which an actor attempts to regulate the behavior of the other, and (2) affiliation, that is, the extent to which on individual regards another positively. According to these au-

[5] The quotation marks indicate that "confidence" in Barry and Crant's view is different from Luhmann's conception, and differs in its meaning from the way it was defined earlier.

thors, social relationships are invariably defined in terms of these two dimensions, because "they are a product of our evolutionary heritage" (Dillard et al. 1996: 706, see also Bugental 2000). At this point, it is not necessary to decide on the dimensionality of relational communications in general. The important point to take here is that individuals "frame" their relationships with the help of an overarching "model" of their relation, which builds on the relational perceptions they have, and changes with the relational communications that occur.

Dillard et al. (1999) use the concept of *relational frames* to indicate those "mental structures of organized knowledge about social relationships ... [which] simplify the problem of interpreting social reality by directing attention to particular behaviors of the other interactant, resolving ambiguities, and guiding inferences" (Dillard et al. 1996: 706). They suggest that relational frames are generic, mutually exclusive, and compete for relative salience during interaction. Likewise, Lindenberg (2000, 2003) posits that trust crucially depends on the salience of a generalized "normative frame" or "solidarity frame" in which hedonic and gain-related goals are "pushed into the background" and opportunistic behavior is "suspended." Importantly, he proposes that relational signals and relational communications are the principle motor behind the stable activation of a relational frame, and behind the changes in frame-salience that occur during interaction.

5.1.4. Framing Relationships

According to Clark and Mills (1979, 1993, 1994), relations can be framed either as "communal," or as an "exchange." These *relationship orientations* fundamentally differ with respect to the basic rules of interaction assumed to govern the social exchange. While exchange relationships follow the principle of "giving or taking one thing in return for another," and therefore invoke a norm of weak reciprocity and allow for the rational consideration of costs and benefits, communal relationships are characterized by a distinct, unconditional concern for the welfare of the other, and follow a norm of mutual responsiveness. Actors then voluntarily provide benefits to one another without mentally accounting for the investments made. Braithwaite (1998) directly adopts and extends this framework to argue that the relationship orientations proposed by Clark and Mill represent different "trust norms" which can serve as a basis for trust by providing different interactional rules, norms, and routines—different "relational frames," so to say. Likewise, Sheppard and Sherman (1998) argue that a number of different "relational forms," which arise from the basic structure of interdependence, are used to frame trust relations, each being associated with a different form of risk the and different means to mitigate them.

Note that we have already uncovered a similar distinction between relatively stable relationship orientations as a basis of trust when we looked at models of trust development, which Lewicki and Bunker (1995) have accurately described as "frame changes" (see chapter 3.1.4).

According to the developmental models, each stage is accompanied by an enrichment of the informational basis on which trust rests, by increased emotional investments, and by a shift to more affect-based forms of trust. Only the last stage of identification-based trust is marked by an attributional shift concerning the other's motivation from external, instrumental (or "exchange") motives to intrinsic (or "communal") motives (Rempel et al. 1985), and a shift from a purely cognitive to a primarily affective basis of trust. The common ground that unites the theoretical perspectives just reviewed is the idea that a relatively stable pattern of relating to the other (the "relational frame" or "relationship orientation") develops for the involved parties, and is situationally activated to define the trust relation and a particular trust problem. We have introduced the concept of a *relational schema* to denote precisely those aspects of stored schematic knowledge which "function as cognitive maps to help [individuals] navigate their social world. These cognitive structures are hypothesized to include images of self and other, along with a script for an expected pattern of interaction" (Baldwin 1992: 462). Thus, the proposed "relationship orientations" and the "relational frames" are synonyms of our concept of a relational schema.

Relational schemata are hierarchically structured (Baldwin 1992, Reis et al. 2000): at the highest level, they describe people and relationships in general. The next level includes exemplars of particular others. The lowest level contains role and situation-specific representations (e.g. "husband-as-father"). Relational schemata are an important class of trust-related knowledge, because most interactions are socially embedded. That is, we often base our choice of a trusting act on the relational schemata applicable to the situations we routinely encounter in everyday life. In a broad sense, even social roles constitute a class of relational schemata, as they structure patterns of relating to other, potentially unfamiliar actors. Likewise, generalized expectations of trustworthiness are presumably embedded in some (higher-level) relational schema—that is, we do not only encode "average" expectations of trustworthiness, but, along with that, broader interactional routines and patterns of relating towards other, potentially unfamiliar persons; we thus maintain relational schemata even for "typical" interaction partners as well.

Furthermore, relational schemata "do not fully specify behavior in any interaction or situation, but they comprise a set of rules that strongly constrain the possibilities and that organize responses to violations of rules" (Fiske 1991: 21). According to Lindenberg (2003), four important aspects are contained within relational schemata: (1) a set of rules about one's own and the other's behavior, (2) expectations about the other's behavior based on these rules, (3) the "surmised" expectations of the interaction partner, and (4) a co-orientation of expectations, meaning that each interaction partner assumes that the same schema is used by the other. In short, "the mental model of a relationship is thus more than just a social norm about how to behave. It minimally also includes descriptive and normative expectations and co-

orientation. It is especially this interlocking of expectations that makes mental models so important for interaction" (Lindenberg 2003: 40). By providing a common frame of reference, relational schemata govern the process of mutual perspective-taking in interpersonal trust relations. In the process, relational communications change the relational perceptions contained within these mental models; this enables the actors to compare the perspective they have adopted towards the relationship with that of the interaction partner.

We began this chapter with the claim that the context of the trust relation is not static, but highly dynamic. As we have seen, it is in fact reproduced in a continuous process of communication by which the actors achieve a convergence on a shared situational definition of the trust problem. The structural coupling of successful elementary units of communication leads to the emergence of the trust relation in a sequential process of social framing. In short, the actors endogenously shape the context of the trust relation and "maintain" or "destroy" it dynamically with each communicative act. Relational communication is of particular importance for the successful establishment and maintenance of a trust relation, because it signals the perspectives which the trustor or trustee adopt towards the trust relation. This information is stored and organized in the form of relational schemata. They delineate the type of schematic knowledge structure by which actors frame relationships in general, and trust relations in particular, and are therefore prime vehicles by which actors can generate a "favorable" definition of the situation in a trust problem.

5.2. Trust and Identity

5.2.1. The Concept of Identity

As pointed out by Baldwin, relational schemata include "images of self and other, along with a script for an expected pattern of interaction" (Baldwin 1992: 462). This points to an important aspect of the framing process which has implications for the development of trust—interpretation does not only concern the meaningful definition of external situational elements (other actors, the meaning of their actions, the status of the relationship, the interpretation of structural, normative and cultural constraints, and so forth), but also concerns the actor's self-concept within that situation. In other words, during social framing, actors reciprocally define both their identities and their social identities.

The concept of *identity*, or *self*, is used in considerably variable ways within psychological research, and has been of major scholarly interest for decades (see Markus & Wurf 1987, Baldwin 1992, Turner et al. 1994, Mischel & Shoda 1995, Brewer & Gardner 1996, Stryker &

Burke 2000, Andersen & Chen 2002, Simon 2004).[6] Despite existing conceptual differences which mainly derive from the particular focus of research, i.e. analysis of internal cognitive processes versus exploration of the influence of social-structural conditions, there is considerable agreement on the major properties and characteristics of the identity concept. Most researchers trace the origin of identity theories back to the early works of James (1890), Cooley (1902) and Mead (1934). James drew a first distinction between the pure ego, or "I," and the empirical self, or "Me," which he further divided into the material, the social, and spiritual self. According to James, a person "has as many different social selves as there are distinct groups of persons about whose opinion he cares" (1890: 282), and these are distinct from his core identity that constitutes the "I." Cooley provided the important insight that identity is primarily shaped and experienced in interaction with others, using the metaphor of a "looking-glass self." Mead further refined the analytical dimensions of identity, emphasized the dynamic character of identity as a product of symbolic interactions mediated by language, and highlighted the influences of social structure and society in shaping identity—in short, "society shapes self shapes social behavior."

These early contributions highlight several fundamentals which resonate in contemporary identity research: (1) identity has a personal and a social dimension, (2) it is dynamic, (3) socially constructed and socially structured; and (4) actors typically have access to multiple identities, which are (5) interrelated to varying degrees (Simon 2004: 25, 46). Adopting a structuralist-interactionist perspective, Stryker (1980) proposes that identities are organized along a "hierarchy" that reflects the institutional structure of the society, and, more concretely, the structure of the individual's life-world. Thus, multiple identities are not an arbitrary product or a matter of individual choice, but stem from the social embeddedness of actors in networks of relationships. In essence, the concept of identity "serves to bridge social structure (society) and social person (self)," as it "mediates between structural forces and the social person's responses" (Simon 2004: 25f.).

On the most general level, identity can be defined as a set of "cognitive generalizations about the self, derived from past experience, that organize and guide the processing of the self-related information contained in the individual's social experiences" (Markus 1977: 64). Generally speaking, we can conceive of an individual's identity as a set of meanings applied to the self in a social situation. Like other forms of typical knowledge, it functions as "interpretive structure that mediates most significant *intrapersonal* processes (information processing, affect, motivation) and a variety of *interpersonal* processes including social perception, choice

[6] Due to a difference in focus between North American and European identity research paradigms, the terms "self" and "identity" have been concurrently used to denote similar concepts and ideas—while the North American psychological tradition prefers the term "self," European identity researchers commonly use the term "identity" (Simon 2004: 26). We will use both terms synonymously and interchangeably.

of situation, partner, and interaction strategy as well as reaction to feedback" (Markus & Wurf 1987: 299, emphasis in original). Identity is in fact akin to a mental schema, and the properties, conditions, and constraints of its activation and use are none other than those which were defined in chapter 4 when we analyzed adaptive rationality from the individual framing perspective. For example, different aspects of identity can be rendered accessible and activated in an automatic or controlled fashion, as a function of cues in the immediate situation (Andersen & Chen 2002).

Since identity is characterized by a high degree of multiplicity and malleability, it is difficult to refer to *the* identity of an individual in the sense of a "monolithic" and unchangeable trait. Instead, researchers usually differentiate between chronically accessible core aspects of the self, which are relatively unresponsive to changes, and other aspects, the accessibility of which depends on motivational and social context variables. Thus, the actual and temporary *working self* is composed of stable core aspects and more a flexible layer of self-aspects tied to the immediate circumstances and the current content of working memory (Markus & Wurf 1987, Mischel & Shoda 1995).[7] This perspective suggests a process-oriented view on identity, emphasizing its dynamic formation in symbolic interaction with others (e.g. Stryker & Statham 1985).

According to Mischel and Shoda (1995), the "structure of personality" cannot be found in some invariant or stable core identity, but in a stable organizing pattern of relationships between the "cognitive-affective units" that generate identities. These units are (1) encodings, that is, stored categories for the self, people, events, and situations (in other words: stored schematic knowledge), (2) expectations and beliefs, (3) affective responses, including physiological reactions, (4) goals and values, and (5) competencies and self-regulatory plans. Individuals differ in the way this cognitive-affective system is organized. In essence, "the basic aspects of personality invariance become visible in the relations between the psychological features of the social world and the individual's distinctive patterns of cognition, affect, and behavior" (ibid. 263), while the "personality state," that is, "the pattern of activation among cognitions and affect at a given time in this system" (ibid. 257), changes with the particular situation. Even more thought-provokingly, Turner et al. (1994) suggest that "the concept of self as a separate mental structure does not seem necessary, because we can assume that any and all cognitive resources—long-term knowledge, implicit theories, cultural beliefs, social representations, and so forth—are recruited, used, and deployed when necessary to create the

[7] "The idea is that not all self-representations or identities that are part of the complete self-concept will be accessible at any one time. The working self-concept of the moment is best viewed as continually active, shifting array of accessible self-knowledge. There is not a fixed static self, but only a current self-concept constructed from one's social experiences. Core aspects of self (self-schemata) may be relatively unresponsive to changes in one's social circumstances and, because of their importance, chronically accessible. Many other self-representations, however, will vary in accessibility depending on the individual's motivational state or on the prevailing social conditions" (Mischel & Shoda 1995: 306).

needed self-category. Rather than a distinction between the activated self and the stored, inactive self, *it is possible to think of the self as the product of the cognitive system at work, as a functional property of the cognitive system as a whole*" (Turner et al. 1994: 459, emphasis in original).

Concerning the categories of self-related knowledge, there is a major distinction between the aspects of personal identity and of social identity (Brewer & Gardner 1996). *Personal identity* comprises all self-related schemata that are exclusively related to the individual. It is the "I" in Mead's sense, marking the differentiated and individuating aspects of identity by which individuals adopt their sense of idiosyncrasy. It refers to "self-categories that define the individual as a unique person in terms of his or her individual differences" (Turner et al. 1994: 454).[8] *Social identity*, on the other hand, refers to those aspects of identity that reflect the self in relation to others, social groups, and broad social categories (the "Me" in Mead's sense). It can be further divided into relational identity and collective identity. While relational identity mirrors the knowledge that pertains to the self in personal relationships with significant others (Andersen & Chen 2002, Chen et al. 2006), collective identity corresponds to "internalizations of norms and characteristics of important reference groups and consists of cognitions about the self that are consistent with that group identification" (Brewer & Gardner 1996: 84). The distinguishing characteristic between these different types of social identity is whether it is an individual-level self-concept ("who am I?," personal and relational identity) or a group-level self-concept ("who are we?," collective identity) that governs cognition (Thoits & Virshup 1997).

According to Chen et al. (2006), *relational identity* reflects who one is in relation to significant others. A relational identity "is (a) self-knowledge that is linked in memory to knowledge about significant others, (b) exists at multiple levels of specificity, (c) is capable of being contextually or chronically activated, and (d) is composed of self-conceptions and a constellation of other self-aspects that characterize the self when relating to significant others" (ibid. 153). Importantly, relational identities also contain "affective material, goals and motives, self-regulatory strategies and behavioral tendencies" (ibid. 154), and can thus trigger emotions and behavioral goals that an individual experiences when relating to a significant other. Relational identities also include social role relationships (employee-employer, worker-coworker, doctor-patient etc.), familial relationships (parent-child, husband-wife), and close personal relationships (friendships and sexual partnerships), and, more generally speaking, all types of role identities. Clearly, they are one major component of a relational schema.

[8] We will subsume under this category those self-references that an individual adopts in terms of an "overarching" self-concept towards the totality of his or her identities (the "Mind" in Mead's sense).

In contrast, *collective identity* entails a more "depersonalized" sense of the self, where individuating differences between individuals step back in favor of a common shared group-identity ("we-identity") that fosters in-group/out-group differentiations; it does not necessarily require personal relationships. This is marked by a "shift towards the perception of self as an interchangeable exemplar of some social category and away from the perception of self as a unique person" (Turner et al. 1987: 50). The two main foci of collective identity research are group memberships ("we, the chess club") and broader social category memberships ("we, the working class"). Collective identification does not require direct contact or exchange with others who share category membership. Rather, the identification of a shared position with others in the social world is primarily "psychological" in nature (Ashmore et al. 2004).

5.2.2. Categorization Processes

Several streams of identity research have been particularly concerned with the premises and consequences of social identification, as exemplified in the theoretical paradigms of "social identity theory" (SIT, Tajfel & Turner 1979) and "self-categorization theory" (SCT, Turner et al. 1987). At the heart of both theories is the process of *categorization*—the application of relevant categorical "prototypes" to streamline social perception. Prototypes describe a "fuzzy set of attributes (perceptions, attitudes, feelings and behaviors) that are related to one another in a meaningful way" (Hogg 2006: 118) and capture similarities and differences between relevant category members and outsiders. In other words, prototypes are mental schemata of social groups and social categories. Consequently, the principles of adaptive rationality which govern activation and use (interplay of opportunity, motivation, accessibility, effort-accuracy tradeoffs, etc.) apply to these structures of knowledge as well—in fact, social cognition has been a major area for the development of dual-process models of cognition and person perception (see Chaiken & Trope 1999, Macrae & Bodenhausen 2000).

If a collective group identity is salient, it functions psychologically "to increase the influence of one's membership in that group on perception and behavior, and/or the influence of another person's identity as a group member on one's impression of and hence behavior towards the person" (Turner et al. 1987: 118). Importantly, the adoption of a social identity fosters *depersonalization;* that is, viewing oneself (or others) as having the attributes of a relevant category, rather than looking for more individuating information. The salience of a social identity is accompanied by a perceptual shift from "me" to "us;" from "him" or "her" to "them." Depersonalization can occur with respect to in-group members, out-group members (in which case it is known as "stereotyping," see Bargh et al. 1996, Wheeler & Petty 2001), and oneself. It involves prototypical, rather than individuating, attributes being used for judging and evaluating the target. The shift from individual to social (relational and collective) identity is often regarded as a "transformation" which has severe affective, cognitive, and motivational consequences. For one, social identification can trigger affective responses and emotional involve-

ment to the social category, commonly experienced as a sense of belonging, closeness, interdependence, and attachment (Ashmore et al. 2004). What is more, goals shift from a personal to a collective level (Brewer & Gardner 1996, De Cremer & van Vugt 1999), and self-interest is not defined at the individual level anymore: "Inclusion with a common social boundary acts to reduce social distance among group members, making it less likely that they will make sharp distinctions between their own and other's welfare" (Brewer & Kramer 1986).

The impulse for the development of SIT and SCT was the empirical observation that even minimal and arbitrary group boundaries can be sufficient to induce in-group/out-group differentiations, and that these have a range of cognitive, affective, and behavioral consequences (Tajfel 1982). According to SIT, (1) humans have a basic need to maintain a positive identity ("self-enhancement"), and this need (2) translates into an implicit drive to create, maintain, and enhance the distinctiveness of groups, whenever the basis of identification changes from a personal to a collective identity—for example, a group membership. In short, when categorizing themselves as group members, the need for positive social identity motivates group members to differentiate their in-group from relevant out-groups. Research in the SIT paradigm has traditionally focused on categories such as gender, race, nationality, and class, and this framework was primarily used for the explanation of intergroup discriminations and conflict.

In contrast, SCT presents a more general theoretical framework, specifying the antecedents and consequences of personal and social identities to explain (inter)individual and (inter)group behavior, as well as the transition from one form of behavior to the other. SCT elaborates on the details of the categorization process as the cognitive basis of group behavior. According to SCT, identity starts with the process of self-categorization, that is, "cognitive groupings of oneself and some class of stimuli as the same ... in contrast to some other class of stimuli" (Turner et al. 1987: 44). Categorization accentuates perceived similarities between stimuli belonging to the same category and differences between stimuli belonging to different categories (i.e. depersonalization). The activation of relevant (self-)schemata during categorization is governed by "relative salience," which is a "function of an interaction between the 'readiness' of a perceiver to use a particular self-category (its relative accessibility) and the 'fit' between category specifications and the stimulus reality to be presented" (Turner et al. 1994: 454).[9] In the terminology of the model of frame-selection, the activation of relevant self-schemata is guided by the match and respective activation weights—we would have to add, however, that the mode of information processing plays a crucial role as well. In fact, the

[9] "Fit has two aspects: comparative fit and normative fit. *Comparative fit* is defined by the principle of meta-contrast … Stated in this form, the principle defines fit in terms of the emergence of a focal category against a contrasting background. *Normative fit* refers *to* the content aspect of *the match between category specifications and the instances being represented*. The interaction between perceiver readiness and fit is assumed to be a general process at work in categorization, not merely one that applies to social and self-categorization" (Turner 1994: 454, emphasis added). This statement of a "logic of appropriateness" was also captured in the MFS by the activation weights and the mode-selection threshold.

idea that a large part of social cognition, especially in the form of "stereotyping," is highly automatic, has attracted considerable attention (see Bargh & Chartrand 1999, Evans 2008). All in all, we can resketch the core tenets of SIT and SCT in terms of individual framing processes—the activation of relevant identity schemata is guided both by accessibility and situational cues indicating the "appropriateness" of a particular frame and script, and the application of an associated (social) identity can occur both in an automatic and controlled fashion.

The social facets of identity—relational and collective—are of particular interest for trust research. Generally speaking, if we conceive of an individual's identity as the set of meanings applied to the self in a social situation, then the question of which identity a trustor assumes in that situation will have critical consequences for the avenues of trust development. More concretely, the affective, cognitive and behavioral consequences of social categorizations can influence the way trustors deal with a trust problem. Unsurprisingly, a number of scholars have proposed that social identities (both relational and collective) can be a basis for trust development (Brewer 1986, Meyerson et al. 1996, Jones & George 1998, McKnight et al. 1998, Burke & Stets 1999, Messick & Kramer 2001, Williams 2001, Tanis & Postmes 2005, Brewer 2008).

For example, Brewer (1986, 2008) argues that a salient collective identity can be a sufficient solution to a trust problem, leading to a form of "depersonalized trust" based on category membership. First, a salient social categorization can be used as a heuristic cue for guiding the activation of relevant cooperative scripts, for example, the reciprocity norm. Second, trustors may project their own attitudes and beliefs onto the group ("false consensus effect", Ross et al. 1977). Thus "assuming that most individuals have generally positive views of the self (high self-esteem), attributing one's own characteristics to others in the in-group will be biased in the direction of positive traits and behaviors, producing a general positivity in thinking about in-groups that is not extended to out-groups" (Brewer 2008: 222). Lastly, the identification with a social group—the activation of a collective identity—may transform individual goals so that, when social identification is strong, goal transformations provide a basis for inferences about the other's favorable trustworthiness. Similarly, Messick and Kramer argue that "when group membership is made salient, a bond may be induced that facilitates trust and mutual aid. Since common group membership characterizes both parties, it induces, in effect, reciprocity" (2001: 101). They emphasize that this type of "category-based trust" (Kramer 1999) does not rely on dyadic embeddedness or on a history of interaction, and can be extended to strangers in situations where a common identity is evoked and provides information about trustworthiness, because the recognition of a shared identity provides a basis for developing expectations about trustee characteristics.

Williams (2001) extends this argument by looking at the impact of social identification on trust in *inter*group relations, asking how dissimilar group memberships affect trust between

in-group and out-group members. According to Williams, the decisive factor influencing how a salient collective identity influences perceived trustworthiness of out-group members is the structural interdependence between the dissimilar groups, which may be cooperative, competitive, or neutral. A competitive interdependence "refers to the perception that an out-group represents a threat to the goals of one's in-group or to one's personal goals" (ibid. 392). Thus, it "may lead to negative category-based perceptions of out-group members' trustworthiness" (ibid.), because both benevolence and integrity of out-group members cannot be confidently assumed. The opposite is true for cooperative interdependence. In addition to this cognitive side, which primarily affects expectations of trustworthiness, she posits that social identification in the presence of dissimilar groups triggers affective responses which reinforce the cognitive consequences of category-based social cognition.

A large body of empirical research supports the general hypothesis of "in-group favoritism" that social identity theory has provided (see Brewer 1979, Messick & Mackie 1989, Brown 2000). With respect to the impact of salient collective identity on trust, empirical data show that various measures of trust and risk-taking are significantly higher when trustees are in-group members, rather than out-group members (Brewer & Kramer 1986, Buchan et al. 2002, Tanis & Postmes 2005, cf. Güth et al. 2008). What is more, the low levels of observed trust towards out-group members can be traced back to the generation of unfavorable expectations of trustworthiness and reciprocation (Tanis and Postmes 2005). An important caveat to this general conclusion was provided by Buchan, Johnson, and Croson (2006). They relativize the general findings by showing that in-group biases in trust depend on the broader cultural context in which interactions take place. In essence, in-group effects are particularly pronounced in individualist cultures, whereas they do not extend to collectivist cultures (see Triandis 1989, 1995). This demonstrates a mediating influence of the cultural "trust settings" (Giddens 1990) and the "culture of trust" prevalent in a society (see chapter 3.2.4).

Burke and Stets (1999) and Tyler (2001) suggest another way by which identity becomes relevant to trust: actors use their associations with groups and organizations to judge their own social status, and through that, their self-esteem and self-worth. Every interaction thus is also a touchstone of "self-verification;"[10] it can lead to the confirmation or negation of one's own personal and social identity. According to Burke and Stets, self-verification is a causal antecedent to trust. They argue that "insofar as a person's identity is verified repeatedly in interaction with others ... that person will gain knowledge of the other's character and will come to trust those specific others" (1999: 351). Moreover, trust through self-verification leads to commitment, emotional attachment, and the development of a shared group orientation. Any

[10] "In self-verification, individuals seek to *confirm* their self-views, often by looking at the responses and views of others… Self-verification involves the cognitive process of matching the self-relevant meanings in a situation to the meanings that define the internal identity standard and guide behavior in a situation" (Burke & Stets 1999: 349).

"mismatch" between the meanings carried in the current "identity standard" and perceptions of corresponding self-relevant meanings in a situation causes an "error signal" which translates into negative subjective experiences (conversely, a reduction of the error signal results in positive feelings). That is, a discrepancy between self-views and the socially expected identity standards triggers negative internal responses. This is particularly pronounced when other actors deliberately communicate that a socially expected identity standard has been violated—actions that indicate a transgression are normally experienced as a deprivation of social approval.

According to Elster (2005), a violation of *moral* norms results in feelings of guilt in the actor, and of anger and indignation in observers, while the violation of *social* norms triggers shame in the actor and contempt in the observer.[11] Note that a breach of trust often taps on both a moral (obligation, benevolence) and a social (reciprocity) norm. Presumably, decisions about trust and trustworthy responses are therefore particularly informative to evaluate both one's own and the other's identity. Trustors use available cues both to assess and learn about the trustee's identity, and to evaluate their own identity in the light of observable responses to their trust. Likewise, trustees learn about the identity of the trustor and use his actions (trust or distrust) to evaluate their own identity. Just as any other social interaction, trust problems offer an opportunity for the "looking-glass selves" to adjust self-conceptions and conceptions of the social identity of the interaction partner.

5.2.3. Signaling Identities

Taking things together, the previous sections suggest that the communication processes involved in the social framing of a trust problem do also supply cues to the identities of the interaction partners. Actors can use these cues to make inferences about the motivation and preferences of the other. The choice of a trusting act rests on a mutual understanding and a temporary acceptance of the identities presented by trustor and trustee. Only if they are perceived as situationally valid can they become a basis for a subsequent interaction and the emergence of a trust relation, and the prolonged continuation of the interaction so initiated (the structural coupling) depends on a sustained acknowledgment of a given self-definition, or its change into another (Henslin 1968, Endress 2002: 55). In other words, the interpretation and acceptance of the identities of trustor and trustee are important events in the process of trust development. The perception of identities is prestructured by the stock of available interpretive schemes in the form of prototypes or stereotypes, and it can be influenced by the "self-

[11] "*Moral norms* include the norm to keep promises, the norm to tell the truth, the norm to help others in distress; and so on. *Social norms* include norms of etiquette, norms of revenge, norms of reciprocity, norms of fairness, norms of equality, and so on. Some norms, about which more later, have features in common with both moral and social norms" (Elster 2005: 202).

presentation" of the actors and their management of the "personal front" (Goffman 1967, see next section).

Bacharach and Gambetta (2001) advance this argument and propose that the signaling of identities is a core process in the development of trust. Their work aims towards a formalization of this perspective in the framework of principal-agent theory. To the *primary* problem of trust, which equates to the choice of a trusting act and thus represents a decision-making problem, they add the *secondary* problem of trust, which must be solved even before the primary choice problem can be considered. The secondary problem of trust is wholly concerned with the credibility of observable signs of trustworthiness ("manifesta") with respect to their power to indicate the nonobservable trust-warranting properties ("t-krypta") of the trustor.[12] Since opportunists can use strategic mimicry to simulate t-krypta, the secondary trust problem can be interpreted as a special case of a signaling game. A signal is "an action by a player (the 'signaler') whose purpose is to raise the probability that another player (the 'receiver') assigns to a certain state of affairs or 'event'" (ibid. 150). Bacharach and Gambetta (henceforth BG) strive to delineate the conditions that need to prevail in order to generate separating equilibria in which manifesta can reliably signal trustworthy types. Generally speaking, separating equilibria exist whenever the costs of using the signal differ between mimics and nonmimics in such a way that it is not profitable for mimics to use them, while it is profitable for trustworthy types.

A normal signaling game would produce an inference structure of the form m → v → t, that is, an inference from manifesta *m* over types *v* to trust-warranting properties *t*. BG add a layer of "identity signaling",[13] so that (g → i) → v → t, whereby identity signals *g* allow for an inference of the social identity *i*, from there to the type *v* and, in this way, an inference about the trust-warranting properties *t*. Identity itself is a krypton, however. It is not directly observable, but can only be signaled. Thus, the trustor principally faces the problem of credibility again, as well as the possibility that identity signals will be strategically exploited to signal a certain type of identity. However, identity signals often have *unique* authenticating characteristics ("signatures") that can hardly be imitated. If a trustee has honored trust in a previous interaction, then the display of his signature can be sufficient to induce favorable expectations of trustworthiness.

[12] "One observes, for instance, physiognomic features—the set of the eyes, a firm chin—and behavioral features—a steady look, relaxed shoulder—and treats them as evidence of an internal disposition. Trust-warranting properties—honesty, benevolence, love of children, low time preference, sect membership—may come variably close to being observable. But, except in limiting cases, they are unobservable, and signs mediate the knowledge of them" (Bacharach & Gambetta 2001: 154).

[13] "Identity signaling is a strategy for signaling a krypton that works by giving evidence of another krypton, that of being a reputation-bearer" (Bacharach & Gambetta 2001: 163).

Signatures are "heteronymous" in that one signature differs from all other signatures allocated by the random action of nature—for example, the face. According to BG, the effectiveness of face-to-face interactions in producing trust is due to the fact that they allow for a cost-free presentation of signatures (facial displays whose recognition is highly automatic) which, on top of that, are also protected against mimicry. But there are more options of producing credible t-manifesta and signatures. For example, trustees who possess t-krypta produce "cues" which are often highly automatic. By definition, these are cost-free to display for those who are trustworthy (honest look, emotional display, voice etc.), and can be used as a credible signal of trustworthiness. Moreover, a "group signature" may allow for identity signaling via some signal of a social identity if the group can establish a reputation or trustworthiness and protect t-manifesta against exploitation ("categorical identity signaling"). Thus, if a trustee signals that he has social identity *g*, and if the trustor has learned that this category has t-krypta, then the display of this categorical social identity can be sufficient to induce trust.[14]

As Bacharach and Gambetta (2001) point out, identity signals are a most relevant aspect of defining a trust problem; they regularly provide credible information about the other's identity, and thus about the trust-warranting qualities of a potential trustee. The authors formulate their argument from a perspective of strategic interaction, and ask which kind of signals can be reliable and credible given that actors are rational. Starting from this assumption of "rational opportunism," they posit that signals are often strategically feigned to initiate and exploit a trust relation, in that opportunistic actors mimic trustworthy types. However, as BG note, the signaling perspective of the primary and secondary trust problem is tied to very stringent assumptions which derive from the rational-choice perspective underlying the principal-agent framework. In short, in order to make inferences of the kind, type, and logic proposed in their model, the fully rational actors would have to know the costs and utility associated with all outcomes, the signal costs and utility for all types, and the probability distributions of the types in the population (ibid. 161). Their model is an example for the strategic interpretation of signals by rational agents under the assumption of full information and rationality. In other words, the object of their analysis is the inference of trustworthiness based on identity signals and the choice of a trusting act in an "ideal-type" rational mode of information processing.

In contrast, the framing perspective of adaptive rationality developed in chapter 4 demonstrates that bounded rationality may lead to automatic trust as well—to an automatic activation of a corresponding trust frame and script by significant symbols and communicative acts which suppresses strategic considerations and which, in the ideal-type case, leads to an un-

[14] Thus, BG are primarily concerned with categorizations of others, or "stereotyping," and not with the impact of self-definitions on trust development. Identity signaling is explored from the perspective of the trustee and conceptualized as the trustee's problem of communicating a trustworthy impression.

conditional choice of a trusting act. This means that the *secondary* problem of trust (which is none other than the problem of interpretation and the definition of the situation), assuming adaptive rationality, can be solved in ways differing from the rational-choice principal-agent perspective. Adaptive rationality applies to sender and the receiver alike, to both the trustor and the trustee. Thus, not only can the interpretation of signals be highly automatic, but also their emission and communication. More concretely, the signaling and interpretation of identities can be controlled or automatic.

Senders will often routinely activate and enact those parts of their social identity which have been identified as relevant in a particular situation, given that a corresponding identity schema exists and is (chronically) accessible (Macrae & Bodenhausen 2000, Andersen & Chen 2002). Likewise, receivers highly routinely categorize the presented social stimuli and cues into available relational or collective categories. In this line, Andersen and Chen (2002) argue that relational identities, by the principle of *transference*, may even be activated in contexts where the particular significant other is not present and where the situational cues are only proximally identic. Likewise, Huang and Murnighan (2010) consider the possibility that relational identities can be activated unconsciously and thereby influence the choice of a trusting act. In short, symbolic interaction and the signaling of identities often occur in an implicit, unintended and automatic fashion, guided by the principles of adaptive rationality.

All in all, the constitution of a trust relation is dependent upon a shared definition of the situation in which trustor and trustee converge on an (implicit or explicit) understanding of the trust problem—a state reached with the help of symbolic interaction and communication during the process of social framing. This does not only refer to the shared understanding of the rules, roles, or routines governing the transaction (i.e., the sources of trust-related knowledge and the "framing" of the relationship), it extends to the proper understanding of the other's identity. Henslin correctly summarizes this idea in saying that, "where an actor has offered a definition of himself and the audience is willing to interact with the actor on the basis of that definition, we are saying trust exists" (Henslin 1968: 140).

5.3. Active Trust Production

5.3.1. Active Trust

The discussion of the secondary trust problem shows an opportunity to create trust actively. If trust is the product of an open and reflexive communication process, are there then possibilities for actors to facilitate its emergence during interaction? Can trust be actively influenced and produced by communicating, choosing, and presenting relevant identities to others, and by managing the impressions generated during interaction? Giddens argues that trust "has to be worked at—the trust of the other has to be won" (Giddens 1991: 96). He introduces the no-

tion of *active trust* (Giddens 1994) to capture the idea that trust is continuously and reflexively reproduced by the actors involved in an ongoing process of social framing. Importantly, the concept of active trust reflects the openness and contingency inherent in trust-related interactions by recognizing the freedom and autonomy of the other, but emphasizes at the same time the power of individual action to influence the other's perspective and to deliberately define the trust problem in a desired way.

The notion of active trust points to a creative element in trust problems, which manifests in the intentional actions and communications that trustor and trustee engage in when interpreting and defining the situation. That is, rather than assuming merely passive trustors and trustees who only draw on their trust-related knowledge to make sense of their perceptions, the parties are directly and actively involved in the construction of a "favorable" perspective towards the trust problem within and beyond the contexts they find themselves in. That is, trust is also an "idiosyncratic accomplishment" (Möllering 2006a: 356) that is actively achieved and influenced by the actors in more or less institutionalized contexts.

According to Möllering (2005a), the active character of trust becomes most visible in a situation of unfamiliarity, in which neither rational grounds (based on the payoff structure) nor institutional grounds (based on taken-for-granted expectations) for trust are present. In this case, actors nevertheless engage in "reflexive familiarization" to actively create the conditions that allow for the trust problem to be solved. This requires the, "to continuously and intensively communicate in order to maintain reflexively the constitution of their social world, including the trust games played in social interaction" (ibid. 28). In fact, the process approach to trust, as suggested by the social framing perspective, is very broad and not limited to such unfamiliar situations only—it extends to *all* situations (of dyadic, network, and institutional embeddedness), because the validity and stability of social structures, their taken-for-grantedness, and the regulative power of social institutions is itself a product of reflexive and continuous communications. In addition to symbolically negotiating the trust relation, actors more or less consciously contribute as well to the emergence and maintenance of the institutional and cultural contexts which enable them to trust.

Lewicki et al. (1998), in detailing the dynamics of trust and distrust in relationships, point to an important implication of the active-trust perspective: a state of trust is always a temporary balance and a fragile "quasi-stationary equilibrium" that is never stable: "balance and consistency depictions may be more accurately represented as single-frame snapshots of a dynamic time-series process, as relationships are transformed through new information that becomes available and is processed and interpreted" (1998: 444). The idea of active trust constitution also reverberates in Zucker's (1986) ideal-type of "process-based" trust. This aspect most directly demonstrates how trust materializes in reflexive social interactions which over time reproduce (and reinforce) the conditions that generate it, but which can change and dete-

riorate at any point. Likewise, Lewis and Weigert (1985a) identify a "feedback process" of trust building, so that "trust appears to be an antecedent to, a consequent of, and an emergent from the processes of social exchange" (ibid. 466)—in other words, a product of reflexive constitution in which the actions of the parties involved shape the final outcome, that is, the constitution of a trust relation.[15]

The notion of active trust also emphasizes the importance of mutual perspective-taking and empathy involved in bringing about a trust relation—to the extent that actors "use or develop *similar* interpretive schemes to define the social situation, the parties will tend to *agree* on their perceptions of the level of trust present in the social situation, so adjustment to each other takes place" (Jones & George 1998: 535, emphasis added). This adjustment process is influenced by the relational communications and the identity signals the trustor and trustee emit. Overall, when thinking about trust from a social framing perspective, the achievement of favorable conditions conducive to trust has to be regarded as a mutual achievement of the parties involved, and the openness and autonomy inherent in communication leaves space for a creative element and for the opportunity to actively shape the definition of the situation for the actors. This opportunity relates to both the trustor and the trustee, each of whom both actively and deliberately influences the perspective of the other.

5.3.2. Impression Management

A paradigm that has been particularly concerned with the behaviors and strategies that actors use to change how they are perceived by others is "impression management" research (Schlenker 1980, Jones & Pittman 1982, Leary & Kowalski 1990, DePaulo 1992). Generally speaking, *impression management* refers to the process by which individuals attempt to control the impressions others form of them; a term closely related is *self-presentation*, which is often used synonymously.[16] Impression management addresses both the motivations and the concrete strategies used by actors to convey a certain impression of them to others. Many, if not all, impression management accounts draw heavily from the work of Goffman (1959, 1967), who developed the concept of *dramaturgic action* to denote the fact that social interactions often resemble the "performance" of actors who, as in the theater, have to give a credible expression of the "character" they embody to the audience on the "front stage" of social life.

---

[15] Lewis & Weigert (1985a) use this argument to propose an "irreducible" element in trust, emerging as a property of the interactions and social exchanges between individuals, so that it "is *not derived* from, *nor reducible* to the psychological states of atomistic individuals" (ibid. 456, emphasis added). However, the social framing perspective developed in this work offers a conception of individual-level decision-making processes which explain the emergence of trust relations, and especially their unconditional character, in a reductive sense of methodological individualism. Even when reflexive constitution is an open and volatile process, the causal antecedents of trust have to be traced back into the psychological—and information-processing—states of individual actors.

[16] As Leary and Kowalski (1990) point out, the term *self-presentation* is, in a strict sense, narrower because impression management may include the management of entities other than the self, and impressions may be managed by means other than self-presentation, for example, by third parties. We will use both terms synonymously here.

Actors thus try to convey a certain impression of their character and their intentions with the dramaturgical means of self-representation. Goffman gave particular attention to the actor's performance of social roles, and the way in which individuals establish a sense of situational normality using interaction rituals to negotiate and maintain the interaction order, that is, the background expectations and the "rules of the game" defining the situation.

What does motivate actors in the first place to manage the impressions they have on others? According to Schlenker (1980), the individual motivation to engage in impression management is subject to expectancy-value tradeoffs: every image that a person might claim has potential benefits and costs, that is, social and material outcomes which differ with respect to their utility and the subjective probabilities pertaining to their successful enactment. Leary and Kowalski (1990) propose that relevant goals, their value to the individual, and perceptions of discrepancy between the "actual" and the "desired" self are the primary determinants of impression motivation. Of course, the social context is a primary source of such motivations, especially when social roles are identified as relevant in a particular context and social identification allows for appropriate self-categorizations in terms of a particular social identity (Goffman 1959). Thus, actors engage in impression management following a logic of appropriateness, and often "tailor their public images to the perceived values and preferences of significant others" (Leary & Kowalksi 1990: 41).

This does not mean, however, that interactions are always subject to fraud and deception. Even if impression management is tactical (that is, occurring in a deliberately controlled fashion), there is a strong intrinsic motivation to convey *accurate* self-images to others. For one, actors value certain aspects of their personality and consistently try to present the positive sides of their character in public (Schlenker 1980, Jones & Pittman 1982). Second, the actual self-concepts constrain the range of potential impressions that actors may try to generate by providing information about the probability that they can successfully instill a false impression and "pull it off," when they claim images that are inconsistent with how they see themselves (Schlenker 1980). As Goffman notes, "an individual who implicitly or explicitly signifies that he has certain social characteristics, ought in fact to be what he claims he is" (1959: 13). Lastly, social norms (e.g. not to lie) and moral norms (e.g. to refrain from deceit) normally deter actors from making claims about themselves that are inconsistent with their self-concepts (Leary & Kowalski 1990). Thus, even when impression management can be tactical, people normally select from their totality of actual social identities those that are most likely to be met with social approval and facilitate current goal-attainment—but rarely, actors build impressions on a completely false identity. On the other hand, there is, of course, always room for mimicry and deception, and for the use of impression management in a deceitful way. The deliberate communication of a false identity can be a means of achieving a desired end, and

the "selective" presentation of the self can be used to conceal inconvenient facets of personality and character.

At the same time, self-presentation can be "overlearned, habitual, and unconscious," so that "people sometimes engage in impression-relevant behavior with little attention" (Leary & Kowalski 1990: 37). From the perspective of adaptive rationality and frame selection, it is apparent that a display of a (social) identity can be highly automatic if the conditions of situational appropriateness, internal availability, and accessibility of a corresponding frame are met. Thus, a tactical element need not always be present in impression management, although most theoretical approaches explicitly address the deliberate aspect of intentional performances to convey a certain picture of one's personality and identity.

The construction of impressions addresses all kinds of "ideas" others can have about an actor. This does not only include personal attributes and characteristics, but also attitudes, moods and emotions, roles, status, physiological states, interests, beliefs, and so on. It can be achieved not only in overt action and verbal communication, but also in stylistic and nonverbal behaviors and in physical appearance (Jones & Pittman 1982, DePaulo 1992). Social cognitive research has provided a good amount of evidence that the formation of impressions—in the sense of adaptive rationality—often occurs rapidly and automatically, and that first impressions may have a long-lasting effect on subsequent judgments (Macrae & Bodenhausen 2000, Bierhoff & Vornefeld 2004). In the following, we will concentrate on those behaviors which are specifically relevant to solving a trust problem. As Luhmann argues, every action potentially creates or destroys trust, and every communication is potential evidence for the trust-related qualities of the individual, and a reason for adjusting trustworthiness expectations. Thus, impression management is relevant in most social interactions, and particularly so in trust relations.

5.3.3. Trust Management Strategies

The discussion of active trust in the preceding section suggests that impression management is an ever-present facet of the communication processes related to the constitution of a trust relation. Since at least two parties—a trustor and a trustee—are involved in its active constitution, we can address the idea of trust-related impression management from either perspective, and ask about the particular actions of trustor and trustee that facilitate the reflexive constitution of "active" trust.

For example, Kramer (2006), focusing on trustors, contrasts two "broad strategies" they can take for coping with social uncertainty in a trust problem. On the one hand, trustors can aim for better discrimination, and selectively engage in transactions only with those who will reciprocate trust. On the other hand, they can engage in behaviors that are "aimed at eliciting trustworthy behavior from others, regardless of *their* prior intentions or motives" (Kramer

2006: 72, emphasis in original). While the first strategy questions the efficacy of discrimination as a strategy for improving the outcomes of a trust problem (Kramer takes explicit notice of the signaling perspective developed by Bacharach and Gambetta), the second approach points to an opportunity for the trustor to actively produce trust. Trustors, according to Kramer, can foster its development in that they (1) encourage trustworthy behavior, (2) reward trustworthy behavior, and (3) signal their unwillingness to be exploited. Collectively, these actions aim at reducing the social uncertainty of the trustee (!) with respect to the personality of the trustor, by solving "his" interpretive problem and defining the situation for the trustee. As Kramer claims, this process is "an important route to trust-building" (ibid. 72). Kramer's approach is, however, relatively exceptional: most trust theorists focus on the role of the trustee in bringing about a trustworthy impression and motivating a trustor to choose a trusting act.

In this line, Beckert (2006) argues that "performative acts" of the trustee which precede the trustor's choice are a primary means of producing the willingness to trust in the situation. The trustee's actions aim at producing an image of trustworthiness and represent an "investment" which he will take so long as the utility derived from realizing the content of the trust relation (i.e. instant gratification, future reciprocal obligations etc.) is higher than the costs incurred. Beckert develops this perspective in direct reference to Goffman (1959) and Bacharach and Gambetta (2001), holding that the performative acts of the trustee aim at resolving the secondary trust problem: "The trust-taker has to succeed in convincing the trust-giver of a definition of the situation that interprets it as cooperative; that is, he has to convince him of his trustworthiness. This 'enticement' of trust depends essentially on the trust-taker's performative self-presentation" (ibid. 324). Applying Goffman's concept of dramaturgic action, he argues that self-presentations "not only have the function of producing the impression of trustworthiness, but they also offer a common definition of the situation that prejudices the trust-giver's action" (ibid.). This argument resembles an earlier one made by Luhmann, who argues that selective self-representations provide the criteria on which to build trust, so that "the foundations of trust in a society are adjusted according to the prospects and conditions of self-presentation and the tactical problems and dangers involved in it" (1979: 40). Thus, a trustworthy trustee (in contrast to an untrustworthy one) "will handle his freedom ... in keeping with his personality—or rather, in keeping with the personality which he has presented and made socially visible" (ibid. 39).

Beckert identifies four "performative strategies" of self-presentation which are conducive to achieving the desired image of trustworthiness. First, trustees can try to increase the commitment of the trustor to the trust relation by creating normative or cognitive barriers to withdrawal. For example, by showing commitment, they may try to induce a reciprocal obligation to place trust in them. Thus, the trustee's advance investment "exercises a subliminal compul-

sion" (2006: 327) to comply with the norm of not disappointing the trustee. Second, trustees can signal a congruence of qualities and characteristics of the trustor, and exploit the fact that similarities in status, group memberships, behavior, and lifestyle translate into higher perceived trustworthiness (Zucker 1986: 70ff., Elsbach 2004); for example, by taking "strategic membership" in similar groups or by communicating similar life-style, clothing, speech, or national or ethnic affiliation. Lastly, trustees can aim at managing the impression of their own characteristics, that is, influence the trustor's assessment of trust-related characteristics such as competence, integrity, benevolence, and predictability. This point was also made by Whitener et al. (1998), who suggest that impressions of trustworthiness can be influenced positively by behavioral consistency, by displaying integrity, sharing control, accurate, open communication, and demonstrating concern. In conclusion, performative strategies of self-presentation aim at producing the "appearance" (Beckert 2006: 328) of trustworthiness; the willingness to trust is developed actively in the situation itself.

Elsbach (2004), in summarizing crossdisciplinary work on factors enhancing perceived trustworthiness, concludes that trustees can use three types of "tactics" to manage their "trustworthiness images": self-presentation behaviors, choice of language and physical appearance. According to Elsbach, the general purpose of all three impression management tactics is to trigger some stereotypical categorization which is associated with a favorable generalized expectation of trustworthiness. For example, by displaying similarities to the trustor, a trustee can be treated as in in-group member. By displaying membership of a reputable group, that is, by categorical identity signaling, trustees can manage to be associated with a stereotypically trustworthy group. The aspects of language and appearance, according to Elsbach, work in the same direction: both can serve to underline the image of a stereotypically trustworthy group a trustee wants to claim membership of. In essence, Elsbach implies that the common denominator of all trust-related impression management is its potential to trigger stereotypically trustworthy categorizations.

However, the direct effects of performance—the "concrete" aspects of communication, such as language characteristics, physical appearance, and nonverbal behavior, seem to go much beyond their influence in amplifying only the desired categorical group membership and in framing a particular social identity. For example, Burgoon et al. (1990) summarize work on the influence of nonverbal behavior on source credibility and persuasion, holding that "nonverbal behaviors carry significant import in impression management judgments" (ibid. 142). Their work provides a detailed insight on how distal vocalic (fluency, quality, pitch, tempo, amplitude), kinesic (eye contact, gaze, the way the body is leaning, smiling, facial pleasantness, expressiveness), and proxemic features (body tension, distance, movement) translate into proximal percepts of immediacy, dominance, and arousal in the perceiver. These are re-

garded as influencing attributes of competence, sociability and integrity, and therefore can be directly related to perceptions of trustworthiness and subsequent persuasion success.

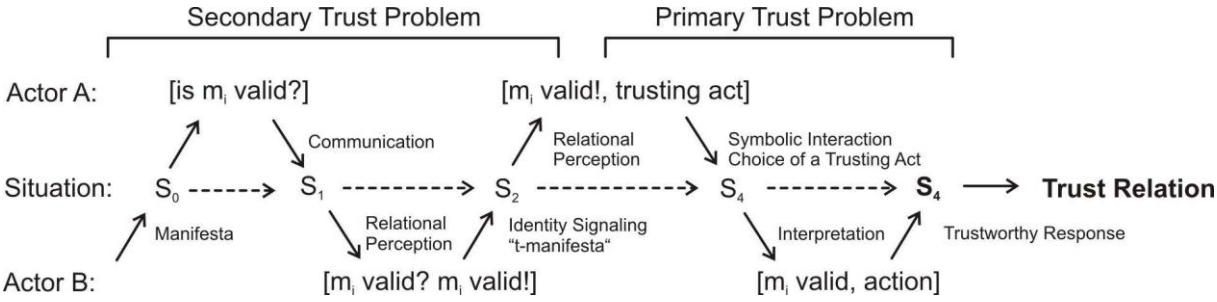
Williams (2007) argues that the most important aspect of trust-related impression management is threat reduction. She claims that "emotional threat regulation"—a process for "managing the harm that others associate with cooperating" (ibid. 596) that involves efforts to influence the emotional responses of others—is a primary means of reducing perceived risks and of inducing trust. She defines threat-reducing behavior as "a set of intentional interpersonal actions intended to minimize or eliminate counterparts' perceptions that one's actions are likely to have a negative impact on their goals, concerns, or well-being" (ibid.)—in effect, these actions represent a cognitive and affective investment into signaling trustworthiness. The potential strategies to achieve threat reduction are (1) altering the situation to remove some or all of the threat-provoking elements, (2) altering attention, that is, distracting the trustor away from the threat-provoking situation, (3) altering the meaning of the situation, that is, "reframing" the facts and critical elements by formulating a plausible narrative that will have a different emotional impact and (4) modulating emotional responses by interrupting a current experience of threat (i.e. physical exercise, alcohol and drugs, relaxing activities). Overall, during threat regulation, the trustee expresses concern, benevolence, support, social competence, and responsibility for the welfare of the trustor. A major consequence of successful threat reduction is the emergence of positive affect and attachment in the trustor and an increase in expectations of trustworthiness. Since threat regulation attempts have a "cost" in terms of interpersonal effort (perspective taking, empathy, understanding, planning) and emotion work, and they represent a credible signal of trustworthiness.

Misztal (2001) and Möllering (2006a,b) directly draw from Goffman in connecting dramaturgic action to the development of trust by focusing on the interactive achievement of situational normality. This achievement depends on the dramaturgic performances of the actors involved and how they manage the impressions they generate to indicate that things are "normal." Generally speaking, by preserving the routine of social life, the actors reinforce the feeling of normality in themselves and others, "which conceals the unpredictability of the reality, thus increasing the perception of general security and trustworthiness" (Misztal 2001: 315). Trust, in this sense, is an "unintended outcome of routine social life" (ibid. 323), and a product of actors who are primarily occupied with enacting a normal social reality in everyday interaction (see chapters 2.3.1 already). Beckert adds to this argument by stating that the operation of institutional mechanisms "cannot be understood independently of the performative production of the willingness to trust" (2006: 329), and without regard for the reflexive nature of the institutional structures themselves, the creation of which is "accomplished in the situation" (ibid.); thus, both situational normality and structural assurance beliefs rest on the en-

actment of corresponding expectations, and on the performances that trustor and trustee take to reassure each other in their mutual intentions.

The following figure summarizes the arguments put forth in the last section, displaying the active social constitution of a trust relation with the help of relational communication, symbolic interaction and identity signaling in a genetic sequence of social framing and communicative acts (figure 19):

Figure 19: Primary and secondary trust problems and the emergence of a trust relation



Overall, the discussion of active trust development, impression management, and the particular trust management strategies that are applicable completes the picture of the social framing perspective put forward thus far by drawing our attention to the concrete performances of the actors involved to create the conditions necessary for a build-up of trust and for the generation of favorable expectations. Impression management research details our understanding of the content of relational communication, and highlights the fact that trust and trustworthiness are active achievements of the actors involved, reached in communication and symbolic interaction during the process of social framing. Both trustor and trustee can proactively take measures to induce a desired response (a trusting act, a trustworthy response) by showing commitment, by reducing perceived threats, or by working on self-presentation of trust-related characteristics. Even the performative creation of situational normality and structural assurance can be addressed under the headnote of active trust creation, highlighting their dynamic and situational character as well as the fragility of these concepts. In sum, the discussion shows that trust has to be understood as an ongoing process of reflexive structuration, in which both trustor and trustee—sometimes implicitly, sometimes explicitly—reach a shared definition of the situation which enables favorable conditions and a confident choice of a trusting act. This pinpoints the last step in the "logic of explanation" of the emergence of a trust relation. Mutual social framing constitutes the building block on which the micro-macro transition and aggregation of trusting choice and trustworthy response into the collective outcome of the "emergent" trust relation occur.

6. Developing an Empirical Test

In the following, the perspective of trust and adaptive rationality which was developed in the previous chapters of this book will be put to an empirical test. This test is has a twofold aim: for one, it is designed to gauge the adequacy of an adaptive rationality perspective in trust research from a general standpoint. The "framing" perspective of trust, as developed in this work, is novel in that it merges psychological ideas of flexible information processing, "situated cognition", and a contingent use of different trusting strategies in trust problems with sociological ideas of a cultural definition of the situation and adaptive rationality. Going beyond previous research, it specifies the causal mechanisms behind these concepts. Adaptive rationality must be regarded as a central dimension of the trust concept, and this demands a focus on the questions of mode-selection, the interplay between the processing-mode determinants, and their causal link to the choice of a trusting act. The following study is designed join these elements in the spotlight of empirical scrutiny.

Second, the test aims for a practical evaluation of the model of frame selection. The model of frame-selection has been used in a number of theoretical and empirical applications in sociological research. For example, the model could be fruitfully applied to model survey response behavior and social desirability (Stocké 2006, 2007b), to explain marital divorce (Esser 1993a, 2002, Hunkler & Kneip 2008), educational aspirations and educational decisions (Stocké 2007a), voter behavior (Kroneberg 2006b), environmental concern and behavior (Best & Kneip 2011), donor behavior (Mayerl 2010), crime causation and criminal behavior (Kroneberg et al. 2010a), the rescue of Jews in World War II (Kroneberg et al. 2010b) and ethnic differences in fertility (Nauck 2010). Taken together, these studies support the major implications of the model of adaptive rationality, even with respect to more "ambitious" interaction hypotheses which are implied by the mode-selection threshold. However, these studies have not been able to confirm the predicted effects with sufficient statistical certainty. The study designs used up to this point (quasi-experimental, ex-post-facto- or survey-based) were limited in their power to draw valid causal conclusions (Opp 2010).

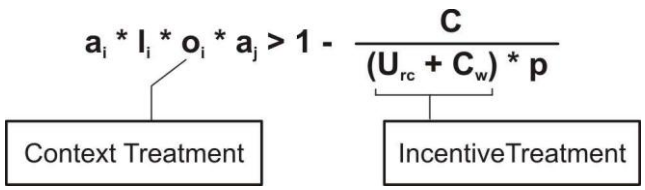
Therefore, this study adopts the method of *controlled laboratory experiments* for the first time. A number of arguments can be brought forth in favor of an experimental approach (see Levitt & List 2007, Falk & Heckman 2009), in particular so when theory is used to model choice behavior. In contrast to field studies, laboratory experiments enable a controlled variation of the decision-making environment. The researcher can manipulate or fix a number of factors which cannot be controlled in a natural setting. For example, he defines the material payoffs and incentive structure, the nature of interactions, the order of interaction and repetition, and the information that the subjects possess when they make a choice. Likewise, institu-

tions, which are normally an endogenous product of social action, can be designed and exogenously manipulated at relatively low cost in an experiment (for example communication, reputation mechanism, contracts and punishment etc.). This amount of control allows for a precise test of hypotheses derived from a theoretical model, and it focuses research on the causal factors of interest; in our case, on the determinants of mode-selection. The general course of action is to operationalize and manipulate the parameters of the mode-selection threshold, while controlling or holding constant all remaining determinants. This provides a causal test of hypotheses addressing the emergence of conditional and unconditional trust.

The experiment demands a proper operationalization of the theoretical constructs and the development of an adequate study design. In using an experimental approach to tackle the question of trust and adaptive rationality, one distinct advantage is that trust research is already equipped with a number of well-established designs which can be fruitfully adapted to the current research question. In the experimental approach to trust, the "investment game" (Berg et al. 1995) is one of the most prominent means to establish a behavioral indicator of trust. It will be extended here with two treatment conditions. Apart from a behavioral measure of trust, the decision times of the participants will be recorded in the choice stage of the experiment as well. Social psychological researchers often use such latency measures to draw inferences about the adopted processing mode. Survey-based measures of trust will be used to operationalize the chronic accessibility of trust related frames and scripts, which constitutes another determinant and parameter of the mode-selection threshold. However, these will be used as independent variables in the analysis.

The experiment focuses on the manipulation of two parameters of the mode-selection threshold: (1) a context treatment varies the presence (or absence) of situational cues indicating the appropriateness of trust-related knowledge and (2) and an incentive treatment varies the initial endowments of the participants to influence the parameter of extrinsic cognitive motivation. Both treatments elicit a direct effect on the parameters of the mode-selection threshold, and therefore influence the choice of a trusting act and corresponding trusting strategies (see figure 20):

Figure 20: Experimental treatments and the mode-selection threshold



The experiment is conducted as a 2x2 between subjects factorial design. Testing the model requires the specification of a set of auxiliary bridge-hypotheses which establish a link be-

tween processing modes, trusting strategies, the experimental treatments, and the observable indicators. While unconditional trusting strategies are expected to result in high levels of trust, conditional trusting strategies support any level between full trust and distrust. With respect to decision times, the automatic mode is expected to be fast, while the rational mode is expected to be comparatively slow. Accordingly, conditional and unconditional trusting strategies are also expected to differ with respect to the recorded decision times.

Although a number of very general propositions have been already put forward in chapter 4, the development of the empirical test requires the derivation of hypotheses which make more precise predictions about the expected statistical effects. As will be shown, the model of adaptive rationality, in conjunction with the set of auxiliary hypotheses, can be used to derive a closed set of admissible interaction patterns that can be expected from a statistical model in the present experiment. These patterns are a specific feature and consequence of the adaptive rationality perspective, in that the mode-selection determinants interact at every stage of frame-, script-, and action selection. From a methodological standpoint, the empirical content of such a model is higher than that of a model which can predict main effects only. The interaction patterns also suggest that any empirical analysis of the trust phenomenon needs to be attentive to the potential heterogeneity in response to the treatments which can be introduced through the interplay of chronic accessibility, situational cues, cognitive motivation and all other mode-selection determinants. The empirical predictions derived in this way pertain to both independent variables, that is, to the choice of a trusting act *and* the corresponding decision times.

The subsequent analysis of trustor behavior rests on the specification of five empirical models which are then tested. The models are applied to both dependent variables in sequence. With respect to the choice of a trusting act, one of the most important results is a confirmed interaction between incentives, the framing of the context, and chronic script accessibility. Previous studies of incentive- and stake size effects have not controlled for the element of chronic accessibility, which is an important mediator of cognitive motivation in the model of adaptive rationality. In line with the predictions generated here, it can be shown that incentive effects do highly depend on the internalization of trust-related scripts (the norm of reciprocity), which counterbalances the negative effects of incentives and high stakes on trust. In other words, trustors who have strongly internalized a trust-related script may be more prone to select unconditional trusting strategies in the face of high stakes than low-accessibility trustors. Similarly, the context is found to influence the choice of a trusting act. If situational cues suggest the validity of trust-related frames and scripts, trustors make use of this information during the choice of a trusting act. Again, the effect of this treatment is found to depend on the accessibility of trust-related knowledge.

In addition, the analysis of decision times reveals a coherent picture. The estimated interaction pattern matches with the analysis of the choice of a trusting act. For example, high chronic accessibility is found to trigger higher levels of trust, but it also leads to a relative decreases in decision time. This suggests a prevalence of unconditional trusting strategies for trustors who have internalized trust-related norms. Moreover, the incentive treatment is found to increase overall decision times, indicating a shift to conditional trusting strategies, but this effect is again mediated by chronic script accessibility. As a result, decision times still are relatively shorter for high accessibility subjects. At the same time, decision times in the context of the trust problem are also strongly dependent on "context free" processing preferences, as measured and controlled for in the form of "faith in intuition" and "need for cognition" scales (Epstein et al. 1996). This is a remarkable finding in itself, as it helps to clarify the tension between intuitive and rational approaches to trust which are an ever-present facet of theorizing. The current data support a perspective of trust in which individual differences in processing preferences crucially shape the mode-selection threshold, the resulting trusting strategies, and the resulting type of trust.

Taking things together, the two behavioral indicators of conditional and unconditional trusting strategies (that is, observed levels of trust *and* corresponding decision times) can be explained with the help of one general theoretical model. The discovery of matching patterns and their similarity over the domain of two different dependent variables suggests the adequacy and validity of the adaptive rationality perspective of trust. The chapter ends with a discussion of the study limitations, potential caveats and highlights questions open to future research.

6.1. Operationalization of Dependent and Independent Variables

6.1.1. The Measurement of Trust

The operationalization and empirical measurement of trust is intricately connected to its theoretical conceptualization. Overall, researchers use three different strategies to quantify and measure trust: (1) a measurement of trust-related attitudes with the help of survey items, (2) an experimental measurement of the behavioral consequences of trust and (3) an assessment with the help of qualitative interviews. A broad conceptualization of trust in the spirit of adaptive rationality suggests that a combined use of different measures is necessary to make precise statements about the types and nuances of trust we encounter. In short, both survey-items and behavioral measures need to be combined if we want to answer how and when conditional and unconditional trusting strategies prevail, and, ideally, such an analysis would be accompanied by qualitative data supplying additional information on the trust development process.

In the psychological literature on trust, a number of scales for the measurement of trust have been constructed and validated (Rotter 1967, Johnson-George & Swap 1982, Yamagishi &

Yamagishi 1994, Couch et al. 1996, Couch & Jones 1997, Glaeser et al. 2000, Fehr et al. 2002b, Kassebaum 2004). These scales are used to assess various aspects of "attitudinal" and "dispositional" trust, such as generalized trust, job- and partner-specific trust, or trust in institutions, social networks, professions and companies. In addition, new techniques to measure implicit components of attitudes ("implicit association test", IAT) have been used to detect implicit aspects of trust-related attitudes. However, as this technique has been developed only very recently, its use is not widespread (Burns et al. 2006, Conner et al. 2007). Most attitudinal and dispositional measures come in the form of a set of survey-items, usually to be rated along a Likert-type response scale. The number of items used per scale varies considerably between the instruments. For example, the GSS (General Social Survey) uses only one item to measure generalized trust ("Generally speaking, would you say that most people can be trusted or that you can´t be too careful in dealing with people?", see Glaeser et al. 2000), while the "Interpersonal Trust Inventory" developed by Kassebaum (2004) involves as much as 55 items. As reported in chapter 3.1, survey-based measures of trust have been empirically connected to a wide range of trust-related phenomena.

Experimental measures of trust in the form of the trust game (TG, Camerer & Weigelt 1988, Dasgupta 1988, Kreps 1990) and the investment game (IG, Berg et al. 1995) establish a behavioral measure of trust. The monetary transfers of the players acting as first-movers are interpreted as an indicator of trust. This course of action is standard in trust research (James 2002b, Hardin 2003), but has recently lead to methodological critique and a consequent refinement of measurement instruments. Importantly, other motivations (such as social preferences, risk and inequality aversion) can easily be confounded with trust. That is why special designs have been devised to separate trust from other influences. For example, Cox (2004) uses a "triadic" design in which first- and second-movers first make decisions *without* a direct counterpart player. This provides a measure that is clean of social preferences, which then can later be controlled for. The approach resembles that of Ashraf et al. (2006), who isolate social preferences with an additional dictator-game measurement. Eckel and Wilson (2004) control for non-social risk aversion with the help of an instrument developed by Holt and Laury (2002). Yamagishi et al. (2005) separate cooperation from trust in a prisoner´s dilemma with variable payoffs, in which the subjects can transform the payoff matrix to their liking (thus, cooperation can occur at low and high levels of payoff interdependence; trust does not manifest in cooperation *per se*, but in the pattern of payoff adjustments that occur over time). Overall, researchers have devised a variety of methods to quantify trust in experiments and control for potential confounding factors that need to be respected.

Concerning the relation between survey- and behavioral measures of trust, Glaeser et al. (2000) did not find any significant correlations, and concluded that survey-items do not assess any attitudes relevant to action at all. Their seminal study provoked a huge body of follow-ups

with mixed results: Ben-Ner and Halldorsson (2010), for example, do find correlations of survey-measures and decisions in investment games, but do not find any influence of risk preferences. Sapienza et al. (2008) and Capra et al. (2008), who both control for social preferences in their studies, do unambiguously find such correlations, while Lazzarini et al. (2003) reproduce the result of the initial study and report none. Capra et al. (2008) can also show, by using a within subject design, that binary and continuous trusting decisions are related. This result is important insofar as it relativizes objections of either game variant as being "inappropriate" to a measurement of trust. In addition, Baren et al. (2010) show that investment game behavior and behavior "in the field" are considerably correlated, providing evidence of external validity of the experimental measures adopted in trust research. Fehr (2008) reports that social preferences are a good predictor for survey-items of trust and concludes that surveys are composed of an expectation-based and a preference-based component of trust attitudes, while experiments provide a preference clean measure. Thus, a combined use of experimental- and survey-based measures is most advisable in research.

An important implication of the model developed in this work is that survey-based attitudinal measures of trust may not exclusively and unconditionally guide the choice of a trusting act. We have encountered a related argument when looking at the psychological development of trust in chapter 3.1 already: it is the "situational strength" of the context which puts a trustor´s general "propensity to trust" in relation to his final intentions (see Gill et al. 2005). Thus, when we ask about the impact of trust-related attitudes and how they can serve as a trust frame, it is important to respect the other determinants of information processing, and particularly, the situational context (cues) and the incentive structure of the trust game (motivation) as well. As will be shown, the model predicts particular interaction effects between attitudinal measures and objective-structural conditions, and the mixed results cited above appear to be indicative of a neglect of relevant variables and incomplete model specifications.

In the present work, survey-based measures will be used as an indicator of the chronic accessibility of trust-related frames and scripts, and thus serve as an *independent variable* in the statistical models. In particular, the chronic accessibility of a generalized trust frame a_i will be measured with the help of a short-version of the "Interpersonal Trust Inventory" (Kassebaum 2004, based on items of Rotter´s ITS). Thus, it is hypothesized that a generalized attitude about unspecific others can serve as an interpretive lens to frame a particular trust relation in the experiment. Furthermore, the chronic accessibility of a trust-related script a_j will be assessed with the help of the "Norm of Reciprocity"-scale (Perugini et al. 2003), which measures how strongly the subjects have internalized the norm of reciprocity (see section 6.1.4. below).

Concerning the measurement of trust and its behavioral consequences, the sequential-game variants of the trust game and the investment game are appropriate for the experimental analy-

sis of the trust phenomenon because they realize most directly the objective-structural conditions as discussed in chapter 2. The investment game subtly differs from the trust game in that decisions about trust are not binary, but continuous. More concretely, when playing an investment game, two players receive some *initial endowment* E. The first-mover (in the role of a trustor) can then decide to transfer any amount X between zero and E from his initial endowment to the second-mover (the trustee). This transfer is regarded as an indicator of trust (see Johnson & Mislin 2011). Importantly, before the trustee receives the transfer, X is multiplied by some factor λ, which represents the surplus and potential gain inherent in the successful establishment of a trust relation and its trustworthy response.[1] After receiving the amount λ*X, the second-mover (in the role of the trustee) can decide to return any amount Y of his total wealth X + λ*X to the trustor. This transfer is regarded as an indicator of trustworthiness. A benevolent trustworthy response requires that the trustee reciprocates with a transfer of at least X, to restore the trustor´s initial wealth. Thus, after both decisions have been made, the trustor A receives a final payoff of:

$$U(A) = E - X + Y$$

Likewise, the trustee B, given $\lambda = 2$, receives:

$$U(B) = X + 2*X - Y$$

If $Y \geq X$, then trust "has paid off" for the trustor and he can in fact realize a utility increase relative to the *status quo* of distrust, which yields safe payoffs E. Otherwise, the trustor experiences a loss and would have been better off to distrust, send a zero amount X and keep E.

For empirical testing and data analysis, we will treat the relative transfer of a trustor in the investment game as a proxy indicator of trust. It is the central dependent variable of the analysis. In order to make results comparable across high- and low initial endowment conditions, we will analyze the transfers relative to the initial endowment. Thus, instead of analyzing the absolute transfer X, the analysis focuses on X/E, the relative amount sent (*reltrust*). The variable *reltrust*, the relative transfer X/E of an experimental subject playing an investment game in the role of the first-mover, serves as an indicator of trust; it is the central dependent variable of the analysis.

[1] The "classical" investment game, as presented by Berg et al. (1995), is played with λ=3. Lenton and Mosley (2011) examine the effect of λ with respect to trust. They hypothesize and empirically find that a large λ positively influence trusting behavior, and that players act more conservative and risk-averse with λ = 2, see section 6.2.1. In the current experiment, an efficiency gain of λ = 2 will be used.

6.1.2. Linking Transfer Decisions and Processing Modes

The model of trust and adaptive rationality developed in this work predicts an (un-)conditional choice of a trusting act as consequence of processing mode selection. One hypothesis that directly relates the prevalent processing mode to observable behavioral outcomes was H8, stating that decision times in the automatic mode should be relatively faster than those in the rational mode. But how can, from the value of the continuous transfers made in the investment game, a conditional or unconditional trusting strategy be classified?

In fact, an empirical classification of trusting choices into conditional and unconditional strategies without the help of fMRI-data proves to be difficult. Arguably, it would be desirable to have a look at "what is going on" in the neural circuitry of the brain while the experiment is run, and thus to have access to data that can be related to modules of the cognitive system known for their role in automatic- versus rational processing. However, such data cannot be collected in the current experiment. Instead, the following *bridge hypothesis* will connect transfer decisions to processing modes. It will be useful to derive concrete statistical predictions with respect to the sign and direction of main- and interaction effects of the model variables when analyzing the transfer decisions as indicators of trust (see section 7.3). In particular, in the ideal-type case, it is assumed that

B1 (unconditional trust): Unconditional trust leads to a complete transfer of resources, X=E.

B2 (conditional trust): Conditional trust supports any transfer between zero and the initial endowment, $X_{\varepsilon}[0, E]$.

B3 (distrust): Distrust leads to a transfer of zero, X=0.

Put differently, *un*conditional trust and a concurrent activation of the automatic mode will lead to higher transfer decisions relative to conditional trusting strategies or distrust, which are triggered by the concurrent activation of the rational mode. There are several arguments that support this assumption. For one, unconditional trust in the automatic mode suppresses the experience of risk and ambiguity in the trust problem. The trustor does subjectively neither question the trustworthiness of the trustee nor consciously process or perceive it. But if there is no perception of risk and vulnerability, if a relevant trust frame can be smoothly activated, a relevant script automatically be used (prompting to a trustful course of action), then there is also no reason to withhold trust, to take precautions and start trust incrementally at a low level. If trustworthiness is taken-for-granted, trustors can confidently expect a benevolent reciprocation, and therefore will unconditionally commit to the trust relation.

In contrast, conditional trusting strategies support precautionary suspicion ("as-if" trust and a "pretense" of suspension). Trustworthiness is not taken-for-granted, the trustor has access to his expectation of trustworthiness, and he also perceives vulnerability, the potential gains and

losses involved in the trust problem, while trying to assess the appropriateness of his initial categorization and judgment of trustworthiness. Of course, the trustor may still arrive at a conclusion in which full trust and a high transfer are confidently selected. This decision results from attributions of trustee characteristics, a consultation of "encapsulated interest", and all categories of trust-related knowledge which cater to a build-up of favorable expectations. But the trustor may as well decide to risk only a relatively small amount, and in fact, any non-zero amount, depending on his expectation of trustworthiness (as, for example, a guilt-aversion model would suggest). This "as-if" trust is conditional and often only mimics real suspension to initiate and test a trust relation. Generally speaking, interventions of the rational cognitive system, on average, should result in lower levels of trust, and hence, in lower transfers as compared to unconditional trust. An empirical demonstration of this effect was provided by Kugler et al. (2009): after experimentally inducing "consequential thinking" among participants in a trust game, the observed levels of trust significantly decreased.

The argument can also be recast in more technical terms: given that information is rationally processed, transfer decisions in the investment game should, on average, approach the Nash-equilibrium. It is for the trustee never to return a positive amount, and for the trustor never to trust and send any positive amount (Berg et al. 1995, Holm & Nystedt 2008). This argument pertains, of course, to situations that involve neither dyadic, nor network or institutional embeddedness. It will be necessary to create an experimental environment in which the subjects do not have a history of repeated interactions, in which there are no reputational mechanisms at work, and in which there exist no explicit institutional mechanisms comforting structural assurance to protect the trustor and sanction a failure of trust. Then, a fully rational actor will never transfer any amount to the trustee. Given that conditional trust *and* distrust depend on an activation of the rational mode of information processing, transfer decisions should approach the Nash equilibrium and be lower than with unconditional trust.

6.1.3. Recording Decision Times

In the current experiment, millisecond time intervals will be automatically recorded at each stage of the experiment (such as reading instructions, answering control questions, making a choice) that provide additional information about the observed choice of a trusting act and underlying processing modes. Generally speaking, decision times (DT) are used both as a dependent or independent variable, and, depending on the research question, they have been used and analyzed as such in cognitive and social psychological research for a long time (see Smith 1968, Luce 1986, Fazio 1990b, Ratcliff et al. 1999, Van Zandt 2000, Mayerl & Urban

2008).[2] In the current work, DT will be used as an *indicator of the processing mode* and the degree of elaboration which a subject adopts during the choice of a trusting act. Using decision time as an indicator of the processing mode is a common procedure adopted by cognitive psychologists, and it has gained increasing popularity in the advent of analyzing and testing dual-process models. Following the general notion of the dual-processing paradigm, the automatic mode is expected to be "fast and effortless", whereas the activation of the rational mode, paired with an increased degree of cognitive elaboration, is expected to be "slow and serial".

This implies measurable differences in the actual time it takes a trustor to decide about the choice of a trusting act. Whereas a short time interval should on average be indicative of the prevalence of the automatic mode and unconditional trust, the opposite holds true for a conditional choice of a trusting act in the rational mode. As stated in propositions 8.1 and 8.2, processing modes should be directly connected to measures of decision time in the present experimental set-up. Therefore, the time it takes a subject from being presented the on-screen decision-making "stimulus" (i.e. the subject is prompted to enter his decision) to confirming the necessary input and making a choice (by clicking a button) will be automatically recorded by the experimental software. As it is, the statistical analysis of DT can provide additional insights about the cognitive processes involved and help us to validate and substantiate the conclusions from an analysis of the relative transfer decision.

One of the main impediments to analyzing decision times is that the "signal-to-noise" ratio is very high (Fazio 1990b). Multiple factors can introduce unwanted variation in decision time data. This noise is not of substantial interest and obscures statistical effects. For example, subjects respond at different rates (that is, they have a different "baseline-speed"), their attention varies from trial to trial, they get confused about a task or question, or they simply forget to confirm an input. Moreover, DT data are typically highly skewed and non-normally distributed; this is partly a result of the presence of extreme outliers from a small but inevitable proportion of respondents who take an extraordinarily long time to complete a task, and partly of the data-generating process itself. Psychologists have long quarreled about the proper way to describe the data-generating process of DT data and how to relate the resulting distributions to cognitive parameters, and albeit a number of different candidates are discussed (i.e. Poisson, ex-Gaussian, Gamma, Wald, Weibull, or Inverse Normal; see Luce 1986, Ratcliff 1993, van Zandt & Ratcliff 1995, Van Zandt 2000), the matter is not settled and researchers use a number of different distributional models and methods to analyze the data. Overall, DT data "can be extraordinarily messy" (Fazio 1990b: 75), asking for close attention to measurement and

[2] Another common term used in psychological research to denote the measure of a time interval between a stimulus onset and a recorded individual response is *response latency*. We will here use the term *decision time* to indicate the conceptual link of this measure to the actual decision, that is, to the choice of a trusting act. Both terms will be used interchangeably here.

data analysis issues. If not taken care of, these issues can distort summary statistics and bias coefficient estimates. A number of statistical procedures have been proposed to deal with decision time data in order to rectify issues arising from (1) the presence of outliers, (2) skewed distributions, and (3) irrelevant noise in the data. It has become a routine procedure to prepare "raw" decision time data into corrected latency measures in an attempt to address these concerns, and to analyze the data using a number of methods which can account for its distributional characteristics.

(1) In order to deal with outliers, decision time data are generally screened for extreme values that can exert a biasing influence on the analysis. Outliers can be identified based on substantial information, that is, when interviewers or experimenters provide information about invalid individual measurements. They can also be identified based on statistical information of the sample, for example, in relation to the standard deviation or some other absolute criterion. Both can be used to define cut-off points and maximum (and/or) minimum acceptable thresholds to identify outliers. A frequent choice is to define a threshold of 2 standard deviations above the arithmetic mean to identify outliers (Bassili & Fletcher 1991, Bassili & Scott 1996). A drawback from such a procedure is that there is no reliable rule as to how establish the cut-offs; their empirical determination highly depends on the sample.

While the identification of outliers is routine, the question of how one deals with them, and whether they should be discarded and assigned as missing or not, is a matter of considerably less agreement among researchers. While some have proposed to impute arithmetic mean values (Stocké 2002) or to replace them with a fixed, pre-defined maximum value (Devine et al. 2002), these techniques necessarily introduces bias into the data, even when they ensure that the number of observations stays constant (Mayerl & Urban 2008: 60). Moreover, using cut-offs can have both advantageous and adverse effects on the power of statistical tests, depending on how precisely the experimental treatments shift the mean and the shape of DT distributions (Ratcliff 1993). Their removal can introduce asymmetric biases into statistics such as the sample mean, median and standard deviation (Ulrich & Miller 1994). In the current analysis, the number of outliers above two times the DT standard deviation from the mean is relatively low (N=9), all models will be re-calculated with and without keeping them in the data set to account for their influence.

(2) Since DT are highly positively skewed, normal OLS models cannot be applied to the raw data, and non-robust measures of central tendency, such as the mean and standard deviation, can be distorted and inflated (Mulligan et al. 2003). A common technique to circumvent this problem is to use data transformations, such as a logarithmic, reciprocal or square root to normalize the data. Each method has a unique normalizing effect on the shape of the distribution and how the "long" DT in the right tail of the distribution are pulled towards the center (for example, logarithmic transformations "normalize" the data stronger, but attenuate the ef-

fect of outliers to a lesser extent than inverse transformations, see Ratcliff 1993). However, a transformation of data necessarily gives rise to interpretation issues. While the ordinal relations of the observations remain intact, the interval and ratio-based relation among the data points is substantially changed in a non-linear fashion, which may distort or even eliminate significant effects.

While the analysis of central tendencies and OLS/ANOVA after a normalization of the data is still a frequent and popular technique, researchers have increasingly used other methods that can accommodate for the overall shape of the distribution (see van Zandt 2000 for a review). The general concern is that the (cumulative) density functions which describe DT markedly differ to that of the normal distribution, and statistical models have to be adjusted to respect this difference. More fundamentally, in cognitive psychology, the distributional forms and their parameters, such as shift, scale and shape, have been directly related to the underlying cognitive processes and architecture in order to derive and justify a certain distribution of DT (Hohle 1965, Townsend & Ashby 1983, Rouder et al. 2003, Matzke & Wagenmakers 2009). An overview and discussion of alternative cumulative density functions which are regularly used to model DT (e.g. Ex-Gaussian, Gamma, Weibull, Lognormal, among others) can be found in van Zandt and Ratcliff (1995), and van Zandt (2000). This enables the examination of treatment effects not only with respect to mean differences, but also with respect to the distribution parameters. A common practice is to fit a certain distribution over the data and interpret changes in the distributions´ parameters as an indicator of treatment effects on cognitive processes (Ratcliff 1978, Matzke & Wagenmakers 2009). In principal, using other distributional forms also enables the fit of a linear model once a proper distribution is specified and accounted for in the statistical model.

A number of authors have used survival models and event history analysis to analyze DT. The observed latency measure is then treated as the outcome of a survival process in which the hazard rate defines the instantaneous propensity to "end" the survival with a response, or a choice, respectively (Box-Steffensmeier & Jones 1997, Johnson 2003, Mulligan et al. 2003). Thinking responses and decision times in terms of hazard rates is a common alternative to investigating distributional forms, because they are less prone to "statistical mimicking" (Luce 1986, Van Zandt & Ratcliff 1995).[3] As Mulligan argues, "the hazard rate fits naturally with how we tend to think about response latency" (2003: 296), and a number of models can be fitted, including for example, non-parametric Cox models (which do not assume a certain dis-

[3] The problem of statistical mimicking describes the fact that distributions with several free parameters are highly flexible, and an empirical DT sample can often be explained by different distributions with equally good fit. If the aim of the researcher is to test underlying cognitive models, and if the predicted distributions, although different, are virtually identical, then DT data cannot be used to discriminate between them.

tributional form), and parametric models in which a specific distribution is specified (such as the Weibull, Exponential, or Lognormal).

To decide about a specific distribution, researchers resort to both theoretical and pragmatic arguments (Dolan et al. 2002). As pointed out above, theoretical models may suggest and generate a specific distribution. Other than that, and from a pragmatic point of view, it is desirable that the distribution provides an adequate description of the empirical shape of the sample. This can also motivate and justify a particular model. In the present analysis, a combination of methods will be adopted to model and analyze DT to provide a check of robustness for the estimated of the effects; these methods will be introduced and discussed in more detail in the chapter on decision time analysis below.

(3) A number of additional factors can increase variability in decision times. Principally, they are regarded as adding "noise" to the data which does not mirror substantial effects of interest, such as subject heterogeneity and measurement error. While the use of computerized software helps to exclude the latter, there is substantial variation in the former aspect in terms of subject heterogeneity. Most importantly, individuals differ in the general speed of responding to an item or task. This difference can be profound and inflate variance in DT data. If individual differences in response latencies are not accounted for when analyzing aggregated data, then treatment effects and between-subject variation are easily confused. Therefore, using within-subjects designs to account and control for an individual *baseline speed* (BS) has become a routine procedure. A common technique to account for this is to use "filler latencies" (Fazio 1990b) and adjust DT measures for the individual baseline-speed of the respondent. Filler latencies are measured on items or tasks which are independent of the target latency. If several measures are used, the baseline speed is constructed by calculating the arithmetic mean of the filler latencies. However, an open question is the precise nature of the filler items (or tasks), that is, their difficulty in comparison to the target task, and their theoretical and thematic closeness (Mayerl & Urban 2008: 64). Baseline speeds can be used in different ways to get a "clean" measure of DT, denoted as DT*. Fazio (1990b) proposes to use either of the following:

(a) Difference Score: $DT^* = DT - BS$
(b) Ratio Index: $DT^* = DT / (DT+BS)$
(c) Z-Score Index: $DT^* = (DT - BS) / SD_{BS}$

In the above formulas, DT is the decision time, BS is the baseline speed, and SD_{BS} is the standard deviation of BS. Thus, difference scores report the observed absolute difference between the DT under scrutiny and the respondent´s baseline speed; it can be negative if a response is faster than the individual baseline. Ratio index scores normalize the observed DT to a range of [0,1], where a value of 0.5 indicates that the DT corresponds to the BS. Important-

ly, ratio index measures accommodate for the fact that absolute differences can stay the same even if the relative magnitude of DT and BS can dramatically differ. Thus, while two observations (DT_1=400 BS_1=200, DT_2=800, BS_2=1000) may have the same difference score (here: 200), the ratio index will be different (RI_1=0.67, RI_2=0.44). That is, in the ratio index, absolute differences between DT and BS are treated as relatively less important with increasing magnitude of the decision time. Lastly, Z-Scores additionally respect the standard deviation of the BS, which necessitates that a number of equal "filler latency" measures have been recorded for each observation. Note that the computation of the scores, as proposed by Fazio, necessitates that a BS can be measured on tasks that are principally identical to the target one, so that the BS and the transformed DT* have a meaningful interpretation.

Mayerl and Urban (2008: 71f.) propose a method to control for an individual BS by estimating the so-called *residual index.* It is derived from a linear regression in which the recorded DT is explained as: E(DT) = a + b*BS + U, that is, as a linear combination of the individual BS, a total sample (task-specific) constant, and residual time U. Note that the residual U can be computed as DT* = U = Y -Y_{hat} = E(DT) – a – b*BS. In other words, by computing the residuals of a linear regression in which DT is regressed on BS, all the variation in DT that is not linearly related to the baseline speed is captured in the residual index U. Positive values indicate that a subject has a longer DT than expected from his baseline, negative values indicate that the response was faster than expected from the baseline. Principally, this procedure is not different to including the baseline speed as a control variable in multiple regressions. Mayerl and Urban (ibid. 77f.) show that the residual index DT*, in contrast to Fazio´s DT*, is not correlated to the baseline speeds after the transformation, while the traditional DT* are still highly correlated to the individual BS.

In the current experiment, latencies will be recorded at each separate stage of the experiment (reading instructions, answering control questions, making a choice). Therefore, there is a stock of tasks which can serve as a filler latency to compute a baseline speed. The time that subjects take to actually decide about the choice of a trusting act will be recorded in milliseconds in the variable *time,* which is the "raw" measure of decision time without baseline speed correction and serves as a second *dependent variable* in the following analysis. A further note on the technical details of the analysis will be given below; a number of different methods will be used to assess the overall validity and robustness of the results. For example, the log-transformed decision times (*logtime*) can be analyzed with robust regression techniques. In that case, it is highly advisable to correct for the respondent´s baseline-speed to get a comparable measure of DT between subjects. A baseline-speed control variable will be computed and introduced. The data will also be analyzed by fitting non-parametric models which make use of the untransformed DT measures to address distributional concerns and accommodate for the non-normal shape of the DT distribution.

6.1.4. Chronic Accessibility of Frames and Scripts

The mode-selection threshold for the unconditional choice of a trusting act was defined as $m_i * a_j > 1 - C / (p * (U_{rc}+C_w))$. A particular important determinant of mode-selection is the chronic accessibility of trust-related frames and scripts which can be applied in the context of the investment game. The activation weight crucially depends on how strongly these mental models are ingrained in the associative memory system and how readily an individual will use them in a situation (see chapter 4 already). During the experiment, the parameters of the threshold will be controlled *or* manipulated with an experimental treatment. To tap on the chronic accessibility of a trust frame and a trust-related script, we will use individual survey ratings and their extremity as a proxy indicator. The rationale for this operationalization stems from a number of results from social- and cognitive psychology research on attitudes, which can be fruitfully combined with the propositions of attitudinal trust research.

Generally speaking, the "strength" of an attitude determines its influence on information processing and behavior, its persistence and resistance to change and persuasion (Krosnick & Petty 1995). Specifically, strong attitudes (1) come to mind faster, (2) persist over time, (3) resist counter-persuasive attempts and (4) guide behavior more than weak attitudes (Petty & Cacioppo 1986, Fazio 1995). However, the concept of *attitude strength* is itself a fuzzy term, subject to an ongoing debate regarding its dimensionality, antecedents and determinants (see Visser et al. 2006 for a review). Attitude strength is regarded as a multi-dimensional concept, measured on roughly a dozen attributes, such as certainty, importance, knowledge, intensity, interest, elaboration, ambivalence, extremity, direct experience, structural consistency and accessibility (Krosnick et al. 1993). Research focuses around the question whether these attributes can be reduced to a single common underlying factor, or whether they represent several unique, or even completely independent dimensions that cannot be combined.

A finding most relevant to our endeavor is that *attitude extremity* and *attitude accessibility* are consistently found to share a common underlying factor, regularly distinct from other dimensions such as importance, knowledge and elaboration (Erber et al. 1995, Pomerantz et al. 1995, Bassili 1996, Visser et al. 2006).[4] Researchers have uncovered positive correlations between attitude accessibility and extremity (Fazio & Williams 1986, Judd et al. 1991), as well as between attitude accessibility and other strength-related attributes listed above, such as importance (Bizer & Krosnick 2001) and involvement (Lavine et al. 2000).[5] Thus, attitudes that are extreme, in the sense of a high agreement or rejection, are also highly accessible; extremi-

[4] *Extremity* is defined as the distance of a rating from the scale midpoint. In the case of a Likert-type scale, the midpoint is the center between the two extremes of the scale in which the respondent rates an item with a "fully agree" or a "fully disagree" statement, respectively.

[5] In fact, almost all pairwise comparisons of the strength-related attributes listed above show such positive correlations (Krosnick & Abelson 1992)

ty may even be a causal antecedent to accessibility (Fazio & Williams 1986, Fazio 1995). In this line, some researchers have used composite indexes combining extremity and accessibility measures into a one-dimensional indicator to investigate attitude properties and processes (Bassili & Roy 1998).

From an intuitive standpoint, the positive correlations between attitude extremity and accessibility make sense: an attitude that we strongly support or reject is most likely one that we can also readily express. That is, "attitudes associated with univalent and extreme underlying structures should occasion relatively little decision conflict and thus should be highly accessible" (Lavine et al. 2000: 81). Importantly, attitude extremity is conceptually rich, as it captures (1) the intensity of feeling an individual experiences with regard to the attitude object, (2) the degree to which an individual holds a qualified position, (3) extent to which a certain attitude or position is regarded as "defendable" and (4) the extent to which an individual would actually defend it (Abelson 1995).

The above findings establish a link between the extremity of an attitude and the latent construct of chronic accessibility. A high rating on a scale gauging generalized trust or the norm of reciprocity, for example, is indicative of high chronic accessibility of a corresponding trust-related frame or script. Simply put, if we do not support the corresponding "trustful" attitude, then the relevant frame or script should also not be chronically accessible to us, and *vice versa*. Survey-based scales for the measurement of trust in effect assume that dispositions to trust, as a relatively stable and persistent trait, can be measured in the same ways as an attitude. Of course, an extreme rejection of the survey items also indicates "accessibility" of some sort. But with respect to the theoretical concerns (the accessibility of frames and scripts that *support* trust and serve as a trust frame) these ratings portray the absence of a corresponding mental model and low chronic accessibility of the trust-related frame or script. In other words, the scale-rating is a proxy indicator of chronic accessibility.

At first glance, this course of action might appear exceptional, given that the most frequently used measure of attitude accessibility is response latency (Fazio 1986, 1990a, 1995). The method we adopt here favors a "meta-judgmental" measure over an "operative" measure (see Bassili 1996). The reasons for this approach are of practical and theoretical nature. First, there is a very practical reason that limits access to latency data: in the course of the experiment, response latencies could not be collected for survey items. The experimental software that was used to conduct the computer-based experiment (z-Tree, see Fischbacher 2007) does not support a measurement of response latencies in the survey-module of the program, even when latencies can be recorded for decision times in the choice-stage.

But there are further arguments that put into question the adequacy of latency measures as an indicator of *chronic* accessibility, suggesting that the approach taken here is more appropriate.

First, on a theoretical level, a resort to the actual scale rating and the underlying substance and meaning of the attitude to the respondent allow for an integration of, and connection to, research focusing on dispositional trust and the influence of inter-individually stable traits. In this area of research, a number of results confirm that the chronic accessibility of trust-related knowledge, in the form of behavioral tendencies and stable dispositions, exerts an influence on a variety of trust-related outcomes (see chapter 3.1 already). Traditionally, these concepts have been measured using scale-ratings. As it is, attitude extremity is the only dimension of attitude strength that is actually related to the content of the attitude and has a "substantial" meaning to it (Visser et al. 2006: 55). In contrast, pure response latencies are devoid of content. They do not tell us unambiguously about trust and the nature of the attitude. If we were to use response latencies without looking at the substantial rating, we would run the risk of attributing high accessibility to both high-trust and low-trust types: both harbor extreme attitudes and will rate the respective scales on their extremes. From the perspective of trust research, the use of actual ratings to assess the chronic accessibility of trust-related knowledge reflects the substantial content of the trust-related attitude, something that latencies cannot capture.

Second, the model of frame-selection suggests that response latencies tell us only indirectly about chronic accessibility, if they do at all. Put sharply, whenever we measure response latencies, a processing mode has already been determined, and both chronic and temporary accessibility have played out their parts in mode-selection. Traditionally, researchers infer accessibility from response latencies, in that a fast judgment points to automatic processing *via* a high accessibility of relevant knowledge. This is why response latencies are also regularly used to directly infer the processing mode (Mayerl & Urban 2008). However, we can never determine whether our measurement taps on a temporary or chronic aspect of accessibility. Accessibility, when measured in response latencies, will represent a fusion of temporary and chronic aspects and mode-selections – they can be influenced, for example, by recent priming and the context. Latency therefore does not capture what the model of frame selection substantially refers to with its concept of chronic accessibility. In fact, the aspect of temporary accessibility is captured in the parameter $a_{j|i}$ and the conditional spreading of activation from a frame to the relevant scripts. What is more, the model suggests that other factors are important during mode-selection as well in that high accessibility alone is not sufficient to guarantee an automatic response under all circumstances (for example, if the motivation for rational elaboration is high). Therefore, using response latencies to tap on *chronic* accessibility is problematic because many other factors (temporary accessibility, motivation, opportunity, context and cues) do influence the processing mode, and therefore latency, as well.

In the present experiment, the chronic accessibility of a trust frame will be operationalized using the individual score of the items of a short version of the "Interpersonal Trust Invento-

ry" (ITI, Kassebaum 2004). This scale measures generalized interpersonal trust towards unspecified others by asking respondents to judge the validity of statements such as "Generally speaking, most people can be trusted" or "You can´t be too careful in dealing with others" (present author´s translation). Since the experiment is conducted anonymously and excludes social embeddedness (no repetition, reputation, punishment) participants cannot make use of other specific categories of trust-related knowledge. Only generalized trust and the relational schema connected to it represent a relevant trust frame in the present experiment. In short, it is expected that participants who score high on the ITI scale can chronically access the trust frame F_i of a generalized-trust relational schema. The resulting normalized score, ranging between [0,1], will be an independent variable (*trustscale*) of the analysis.

To operationalize the chronic accessibility of a trust-related script, we will use the "norm of reciprocity"-scale (Perugini et al. 2003). Reciprocity norms are a primary social mechanism for the control and protection of trust. Importantly, the reciprocity norm has a direct relevance for action, because it suggests that a trustworthy response can be favorably expected, thereby motivating the choice of a trusting act (this recasts A2: $a_{k|j}=1$). We assume that a trust-related frame, if adopted, points towards reciprocity norms as a part of the "rules of the game". If a frame of generalized trust is adopted, the reciprocity norm should be temporarily accessible (this recasts A1: $a_{j|i}=1$). The resulting normalized score of the "norm of reciprocity scale", ranging between [0,1], will be another independent variable (*recscale*) of the analysis.

A documentation of all items used, factor analyses and reliability measures can be found in Appendix B. The individual scale ratings will be constructed by summing up and averaging the scores of the 7-point Likert-type items which could be answered ranging from "I fully disagree" to "I fully agree", leaving open a non-response option ("I don't know") at every item. These measures serve as important independent variables of the statistical analysis. They will be coded such that higher values indicate a higher degree of agreement towards the statements, and thus, higher chronic accessibility.

6.1.5. Intuition and the "Need for Cognition"

Throughout this work, a recurrent theme in the discussion of the trust concept was the idea that trust can be based on different cognitive "routes" and processing modes. It ranges from a rational decision based on the controlled and elaborate "bottom-up" integration of relevant information to a "top-down" use of cognitive short-cuts as a basis for a leap of faith. As we know, the automatic mode is characterized as "fast, effortless, associative, implicit, slow-learning and emotional" (Kahnemann 2003: 698). Answers provided by the automatic route and the associative cognitive system just "pop" into the head and do not provide much justification other than *intuition*. They become part of the stimulus information, rather than being seen as part of the perceiver´s own evaluation or interpretation (Smith & DeCoster 2000).

At the same time, researchers have compiled a large body of empirical evidence revealing that individuals differ in their disposition to actually follow their intuition and to rely on automatic *versus* rational processing in a variety of tasks and judgment domains (see Cacioppo et al. 1996, Epstein et al. 1996). These findings suggest that there exist stable differences in the chronic tendency to activate a certain processing mode; individuals have a "preference" for processing, that is to say. It manifests as a stable, intrinsic readiness to engage in effortful and elaborated thinking, and it results in a corresponding "thinking style."

The most frequent scale-based measure to assess such individual differences is the "Need for Cognition"-scale (NFC), initially developed by Cacioppo and Petty (1982). This instrument captures the "tendency for an individual to engage in and enjoy thinking" (ibid. 116), and reflects the aspect of cognitive motivation that individuals have towards elaborate processing. According to Cacioppo et al. (1996), individuals high in NFC act as highly motivated "cognizers," in contrast to the "cognitive misers" at the low end of the scale. Differences in the need for cognition derive from past experience and behavioral histories, and they influence the acquisition and processing of information relevant to judgment and choice. Respondents high in NFC enjoy thinking, get intrinsic rewards from effortful mental exercises and prefer to confront demanding cognitive tasks instead of easy ones. In contrast, low NFC individuals dislike expending mental effort and try to avoid situations that demand it.

Accordingly, need for cognition consistently influences a variety of judgments, tasks and decisions. Individuals high in NFC are more readily influenced by the quality of persuasive arguments, show better recall and performance on a variety of cognitive tasks, actively search for more information, and are more likely to base judgments on empirical information and rational considerations (see Cacioppo et al. 1996 for an extensive review). Low NFC individuals are more prone to use automatic associations and stereotypes in judgment (Florack et al. 2001), and they more readily use situational cues as a quick-step to interpretation and choice (Smith & Levin 1996, Shiloh et al. 2002). In contrast, high NFC individuals are more resistant to attempts to change reference points through peripheral cues, and overall they are less susceptible to framing effects (Smith & Levin 1996).

In a critical examination of Cacioppo and Petty´s instrument, Epstein et al. (1996) argue that a low motivation to process information systemically ("being a cognitive miser") need not necessarily translate into a high motivation to process intuitively, and *vice versa*. That is, preferences for intuition and deliberation may be independent. They develop a second scale, "Faith in Intuition" (FI), to complement the NFC instrument, and measure both rational and experiential processing preferences with the two resulting unipolar scales (the "Rational-Experiential Inventory", REI). FI captures the "engagement and confidence in one´s intuitive abilities" (ibid. 392) and mirrors the extent to which individuals chronically rely on intuitive judgments. Their study shows that the two constructs are relatively independent, but describe

interindividual differences which are correlated to a number of personality constructs (see also Keller et al. 2000, Betsch 2004). Shiloh et al. (2002) can show that specific combinations of FI/NFC and the resulting cognitive types react differently to framing treatments. They also differ in the extent to which responses in statistical reasoning tasks are based on intuition or deliberation.

The findings presented above have direct implications for trust research. If individual thinking styles and the extent to which actors rely on intuitive or rational processing vary systematically, then the process of trust development potentially varies based on cognitive types. More pointedly, some subjects will be more prone than others to select conditional *versus* unconditional trusting strategies (over and above the differences captured by the chronic accessibility of trust-related frames and scripts and the experimental treatments), simply because they routinely prefer a more automatic or controlled style of thinking about a trust problem. Therefore, it is advisable to explore individual differences in the chronic tendency to rely on intuition and deliberation in the experiment. When analyzing the experimental data, it is important to keep track of such differences in cognitive style: they represent a source of variation that is not attributable to the experimental treatments or the variation in accessibility alone. If the subjects differ in the way they tend to chronically engage a particular processing mode, this will increase within-group heterogeneity and potentially lead to more heterogeneous treatment effects.

To explore the impact of processing styles on interpersonal trust, subjects will be asked a German version of the "Rational-Experiential Inventory" (REI, Epstein et al. 1996) with its subscales "Need for Cognition" and "Faith in Intuition" (see Keller et al. 2000). Appendix B lists the full set of items used in both scales, which were assessed using a 7-point Likert-type scale which also left an "I don't know"-option open at every item. The normalized scores of both measures (*fiscale, nfcscale*), ranging between [0, 1], serve as *independent variables* in the following analysis. They will be coded such that higher values indicate a higher degree of agreement towards the statements, and thus, indicate that the cognitive style in question is agreed and featured by the subject´s self-report.

6.1.6. Control Variables

In addition to the independent variables discussed above, a number of control variables will be collected to gather additional information about the subjects. Generally speaking, trust researchers have uncovered a variety of factors which influence both attitudinal and behavioral measures of trust. A control and analysis of these variables can detail the picture and sharpen our knowledge about the effects of the various experimental treatments and measures, their differential effects on the mode-selection threshold, and the observable consequences for trust.

As it is, the influence of the individual socio-economic background on expectation formation and choice in trust problems has been demonstrated in surveys and experiments (Fehr et al. 2002b, Bohnet & Zeckhauser 2004, Gächter et al. 2004, Güth et al. 2005, Ashraf et al. 2006, Schechter 2007, Capra et al. 2008, cf. Gächter & Thöni 2004). Moreover, the socio-economic background influences individual social preferences, which are regarded as being culturally heterogeneous (Buchan et al. 2002, Cook et al. 2005, Buchan et al. 2006). Researchers have found that a large set of individual attributes has a moderating influence on trust, including gender (Buchan et al. 2008), age (Garbarino & Slonim 2009), ethnicity (Ben-Ner & Halldorsson 2010), familial ties (Ermisch & Gambetta 2010) and religiosity (Tan & Vogel 2008). In particular, the following additional information about the participants of the experiment will be collected:

(1) Age (*age*): Several studies have reported age-effects on a range of trust-related measures. For example, Naef and Schupp (2009) find weakly significant *negative* effects on a transfer decision in an investment game, and Dohmen et al. (2011) show that older subjects are more risk averse than younger subjects. Garbarino and Slonim (2009) demonstrate that age decreases the "sensitivity" to trust for females – reciprocal behavior diminishes as age increases. In contrast, Gächter and Thöni (2004) find that older people trust relatively more in that they have a more *positive* opinion about other´s fairness and helpfulness. However, this does not translate into empirically different measures of generalized trust, trusting behavior and trustworthiness. In contrast, Ermisch et al. (2009) find that age significantly increases trust in a trust game. Overall, age effects may be present in the data, but their direction is not clear. To explore this issue further, respondent´s age will be recorded.

(2) Gender (*sex*): A number of studies have consistently demonstrated that gender differences in trust and reciprocity exist (see Garbarino & Slonim 2009 for an overview). In an early study, Orbell et al. (1994) showed that females are generally expected to be more cooperative. The authors did not find any influence on actual trusting behavior and concluded that generalized role expectations are not of practical matter in particular exchange contexts. But since then, a number of other studies have demonstrated stable gender differences in trusting and reciprocal behavior in experimental trust settings (Croson & Buchan 1999, Chaudhuri & Gangadharan 2002, Cox 2002, Buchan et al. 2008). A consistent pattern that has emerged is that females are more reciprocal, while they are also more risk averse and trust less. Hence, to get a grip on gender differences, we will collect information on participant´s gender.

(3) Relationship status (*partner*) and relationship length (*partner_l*): These variables will be included because the presence or absence of a relationship potentially influences the availability of relational schemata and trust-related frames and scripts. Thus, differences in the accessibility of trust-related frames and scripts are expected depending on whether an individual is actively engaged in a relationship: being in a relationship potentially increases the accessibil-

ity of trust related knowledge. But there may also be factors at work which are counter-productive to trust: Ermisch et al. (2009) find that divorced or separated individuals are *more* trusting than engaged or married counterparts. They speculate that these actors might have a greater incentive for interaction with strangers. Similarly, Ermisch and Gambetta (2010) find that strong family ties prevent trust towards strangers as measured in a trust game, and conjecture that strong family ties prevent outward exposure and the development of sufficiently positive generalized expectations. Thus, being in a relationship may as well decrease levels of trust. To explore these effects, both relationship status and length will be recorded.

6.2. Experimental Design and Method

6.2.1. Experimental Design

In the empirical test, an experiment will be used to manipulate the parameters of the mode-selection threshold. The design involves two treatments which influence the mode-selection threshold in the experimental setting of the investment game: (1) a *context treatment*, which varies the presence of situational objects indicating the appropriateness of trust-related knowledge and (2) an *incentive treatment*, which varies the initial endowments of the participants in order to influence the motivation-component of the mode-selection threshold. The two factors will be varied on two levels. Thus, the resulting experiment is conducted as a randomized 2x2 between-subjects factorial design (table 3):

Table 3: Experimental treatment groups and factor levels

Treatment / Level		Context	
		Neutral	Cooperative
Incentives	High (40€)	High/Neutral	High/Cooperative
	Low (7€)	Low/Neutral	Low/Cooperative

In the version of the investment game adopted here, participants will be facing a multiplier and efficiency gain of λ=2. As Lenton and Mosley (2011) argue, multiplier effects "incentivize" trust because the average expected returns increase. For example, if trustees return half their gain with an average probability of 0.5, then a multiplier of λ = 4 implies that a transfer of 10€ from player A gives player B 40€ and 10€ would be returned on average. In contrast, a multiplier of λ=2 implies that a transfer of 10€ from player A gives player B 20€ but only 5€ would be returned on average. In the experiment conducted by Lenton and Mosley (2011), participants transferred significantly more when efficiency gains were high (λ=4) as compared to low efficiency gains (λ=2 or λ=3). Put differently, if the trustee expects a certain return and compensation for his trust, then expectations of trustworthiness must become increasingly positive the lower is the multiplier λ in order to induce an equal-sized transfer. In fact, several

other authors have used an efficiency gain of λ=2 without explicitly taking into account the incentivizing effects of its modification or providing an explanation of their motivation to do so (Glaeser et al. 2000, Lazzarini et al. 2003, Naef & Schupp 2009).

In the present experiment, the use of λ=2 has a practical and theoretical motivation: theoretically, it is expected to create the most risky environment. Higher efficiency gains may invite "faulty gambling" and convince even distrusting subjects to take a risk and transfer a small amount. In contrast, the λ=2 setting is more risky and therefore should not invite subjects to "just go for it". Trust is more risky because, when holding the expected returns constant, trust needs to be built on an overall more positive expectation of trustworthiness. In other words, to induce an equal transfer in the trustor, expectations need to be more optimistic under λ=2 than under λ>2. Therefore, such a design can better discriminate between subjects that generate favorable *versus* unfavorable expectations – only those trustors with a highly favorable expectation of trustworthiness will be motivated to choose a trusting act. On the practical side, high-incentive treatments are also very costly. If paired with a high multiplier, the costs of the experiment rise exorbitantly, and λ=2 represents the more economical alternative.

The two experimental factors, in addition to measures of frame- and script-accessibility, are the main *independent variables* of the statistical analysis. The experimental setting of an anonymous one-shot interaction in the investment game between the randomized, randomly matched participants serves as the tool to collect the data. In addition to the decisions about trust, the individual decision times (that is, the time spent to make a choice) of the participants are recorded and analyzed to gain more insights about the processing modes and trusting decisions.

To maximize observations and available data, the investment game will be implemented in the following way: all participants will first make a decision in the role of player A (the trustor), then they will be informed about a restart and second round, and then make a decision in the role of player B (the trustee), in which the transfer decision X of a randomly selected and matched participant will be used to determine their total income when deciding about the reciprocal response Y. This is to keep any potential confound that comes from first making a decision as a reciprocator out of the data. However, this course of action has a cost to it: the second-mover decisions of player´s B may be influenced and confounded with the preceding stages of the game. Therefore, an analysis second-mover decisions will not be conducted.

6.2.2. Context Treatment

The context treatment consists of a change in the wording of instructions that are presented to the participants during the experiment. The goal of the context treatment is to (1) decrease the

perceived social distance between participants,[6] (2) create a salient social group identity and (3) alter the "situational strength" of the environment into which the investment game is embedded towards a heightened appropriateness of relevant trust-related knowledge. In particular, presenting and priming the participants with cues that point to the appropriateness of a shared social identity, communal relationship orientations and the appropriateness of trust-related norms increases the temporary accessibility of relevant frames and scripts; for example, the reciprocity norm. In effect, a successful manipulation increases the "match" of trust-related knowledge that can be used to favorable define the situation and shifts the mode-selection threshold towards an activation of the automatic mode.

The manipulation of the presented context as a means to change the definition of the situation of the subjects is a common experimental technique. For example, Burnham et al. (2000) have used the labels "opponent" *versus* "partner" in the instructions of an extensive form trust game to explore the effect of the "friend-or-foe"-heuristic (FOF). Using a "partner" wording to describe the experiment resulted in a significant increase of trust and trustworthiness in the cooperative condition. This effect was initially explained in terms of the activation of the FOF (see chapter 4.1 already). In the more general approach adopted here, these effects are explained as a framing manipulation that affects interpretation and choice and the relational orientation subjects adopt by influencing the temporary accessibility of trust-related knowledge. This is encouraged by the presence of corresponding situational cues and a (potential) shift of processing modes. A closely related framing-manipulation is the use of instructional labels to shift the definition of the situation by changing the "name of the game" (see Ross & Ward 1996, Kay & Ross 2003, Liberman et al. 2004, Dufwenberg et al. 2011). Naming an experiment a "Wall-Street Game" (exchange relationship-orientation) as opposed to a "Community Game" (communal relationship-orientation) significantly influences the cooperation rates observed in subsequent Prisoner´s Dilemma rounds. What is more, even the mere presence of corresponding visual cues (symbol of a bank note on-screen, a business suitcase in the room etc.) can produce comparable effects (Kay et al. 2004, Vohs et al. 2006).

In a similar fashion, Hoffman et al. (2008) show that proposer behavior in dictator games varies with social distance – increased anonymity decreases the distribution of offers. They show that subtle changes in the wording and formulations indicating closeness, community of sharing and the existence of a social exchange framework trigger greater reciprocity (more generous proposal) behavior. In other words, contexts that indicate the relevance of social norms or promote social identification are readily interpreted by individuals, and this information is

[6] Hoffman, McCabe and Smith define social distance as "the *degree of reciprocity* that people believe is *inherent* within a *social interaction*. The greater the social distance, or isolation, between a person and others, the weaker is the scope for reciprocal relations" (2008: 429, emphasis added). Thus, social distance is conceptually directly related to reciprocity and "communal" relationship orientations in which reciprocity is not only expected, but even normatively demanded.

used during the definition of the situation and the choice of action. Likewise, Buchan et al. (2002) manipulate the perceived social distance by implementing a minimal-group design in the investment game using differently colored instructions and wording, or color-coded groups plus personal *versus* impersonal group-discussions (Buchan et al. 2006) and observe comparable (but weaker) effects. From the perspective of social-identity theory, the use of labels that indicate a de-personalized self ("we", "team", or "us" as the point of self-reference) has also been demonstrated to induce shifts from personal to collective selves and corresponding shifts in motivation and expectations (Brewer & Gardner 1996, Tanis & Postmes 2005, see chapter 5.2.2. already). Some scholars have observed weak to none effects of mere labeling treatments, but nevertheless find that payoff-interdependence (Eckel & Grossman 2005, Güth et al. 2008) or a group-building phase with a joint task (Bauernschuster et al. 2010) foster in-group cohesion and result in the creation of a common social identity, group-related goals and motivations in trust settings. Overall, the manipulation of the social distance between experimental participants and the context of the experiment does elicit considerable effects.

In this experiment, the approaches reviewed above are combined to create a context treatment. First, when reading general and specific instructions about the experiment, the participants are confronted with word-pairs that either point to a *neutral*, or a *cooperative* scenario. In the neutral condition, participants are informed that they will be randomly matched into a "group" together with one more "participant." In the cooperative condition, they are informed that they will be matched into a "team" with a randomly selected "partner." These word-pairs are then used throughout the experiment whenever further instructions are presented and a corresponding reference has to be made (the full set of written and on-screen instructions which was used can be found in Appendix B). In addition, subjects will be shown a different welcome screen when entering their computer booth. In the cooperative condition, the welcome screen will show a picture of "shaking hands," in the neutral condition, the participants will be presented a picture of "bank notes" (see Appendix B). These visual primes are used to assist the word-pair manipulation.

The cooperative "partner/team"-condition aims at changing the participants´ definition of the situation into the direction of a favorable interpretation of the trust problem, relative to the neutral condition. Even when there is no direct identity signaling involved, the treatment is intended to change the relational perception and relational orientation of the participants. In contrast to playing with an anonymous participant in a group (potentially involved in an exchange relationship), being a team-member who is working with a "partner" should activate a more communal relational orientation, promote the creation of a shared social identity and point towards the appropriateness of trust-related communal norms, such as the norm of reci-

procity. Overall, the goal of the treatment is to increase the match m_i of a trust-related frame with which to favorably define the trust problem by presenting relevant situational cues.

Thus, it is hypothesized that the parameter o_i increases in the cooperative condition, positively influencing the match m_i. Using bridge hypotheses B1-B3, a direct empirical hypothesis concerning the level of trust and expected transfers X can be derived. Transfers are expected to increase in the cooperative condition. Moreover, additional hypotheses about the expectation of trustworthiness, the influence of instrumental variables and the expected decision times can be generated using the model propositions P6-P8 (see section 6.3). In particular, in the cooperative context, expectations should become more favorable, the influence of instrumental variables should decrease and decision times should be relatively faster (hypotheses will be fully developed in chapter 6.3).

6.2.3. Incentive Treatment

The second treatment manipulates the incentive structure of the trust problem with the aim of increasing the motivation to engage in the rational mode. As it is, motivation comprises both an individual-intrinsic and a situational-extrinsic dimension. While the need for cognition and faith in intuition measures of the REI scale, as discussed in section 6.1.5, capture a stable inter-individual difference in cognitive motivation, the incentive treatment is designed to change the extrinsic component thereof. In the mode-selection threshold, motivation is represented by the composite term (U_{rc}+ C_f). As suggested by the model of trust and adaptive rationality, an increase in motivation increases the right-hand side of the mode-selection threshold, and therefore decreases the probability of an unconditional choice of a trusting act in the automatic mode (H5); we can easily derive corollary hypotheses concerning the effect of monetary incentives on the level of trust, expectations and decision times.

The effect of monetary incentives on judgment and decision-making has been intensively explored in economic and psychological research. However, these studies have revealed inconsistent results (see Camerer & Hogarth 1999, Hertwig & Ortmann 2001). In conducting a comprehensive meta-analysis, Camerer and Hogarth conclude that "in many tasks incentives do not matter, presumably because there is sufficient intrinsic motivation to perform well, or additional effort does not matter because the task is too hard or has a flat payoff frontier" (1999: 8). Incentives usually do *not* matter when the returns to additional cognitive effort are very low (a "floor" effect), or when it is very hard to improve performance with additional effort (a "ceiling" effect). However, even when there is no significant main effect, incentives often alter the variation in the data. Incentive effects are most pronounced in judgment and decision, problem-solving or memory/recall tasks. In a related meta-study, Hertwig and Ortmann conclude that "although payments do not guarantee optimal decisions, in many cases they bring decisions closer to the predictions of normative models" (2001: 395).

For example, Wright and Anderson (1989) hypothesized and empirically corroborated that the use of heuristics, such as anchoring and adjustment, decreases with performance-contingent incentives because they "increase a subject´s motivation, causing the individual to think more carefully and completely about the judgment situation and his or her judgment, and therefore not display any anchoring effects" (ibid. 69). In other words, incentives increased the motivation to engage the rational mode. As expected, the anchoring effect was significantly diminished by the availability of performance-contingent incentives. Likewise, Levin et al. (1988) empirically could prove that framing effects were deleted in high-stake conditions, pointing towards the prevalence of a more controlled information processing state in which context cue validity is scrutinized for "appropriateness" when motivation is high.

Related results have been obtained in the domain of economic research on risk aversion. The general conclusion that can be drawn from a large number of lottery experiments over incentivized outcomes is that risk aversion increases with stake size. Individuals become more risk averse when "much is at stake" and the motivation to carefully consider a decision problem is high (Holt & Laury 2002, Bruhin et al. 2010). Specifically, these effects seem to emerge from a more pessimistic (or: "realistic") expectation-formation over small-probability outcomes, which are normally overestimated (Bruhin et al. 2010). A number of behavioral experiments have explored the effect of stake-size on outcomes. In ultimatum games, it is often found that minimal acceptable offers are lower with high stakes than with low stakes (Hoffman et al. 1996, Slonim & Roth 1998, Cameron 1999). None to only weakly significant effects have been found in the dictator game (Diekmann 2004, List & Cherry 2008), in public-good games (Kocher et al. 2008) and in the gift-exchange game (Fehr et al. 2002a); these experiments have not provided clear and unambiguous evidence that incentives shift decisions towards equilibrium predictions. However, it is important to note that most of the experiments uncover a change in the overall distribution of outcomes, even when mean-comparisons are insignificant (Camerer & Hogarth 1999). This is important insofar as it suggests a heterogeneous response to incentive treatments.[7]

The hypothesis of an incentive effect can be directly derived from the model of frame-selection: an increase in motivation increases the probability of a frame- or script-selection in the rational mode. As a result, a more elaborated and controlled processing of information will prevail during interpretation and choice. However, incentive effects can be counter-balanced by high temporary and chronic accessibility. A high match between situational cues and

[7] As the model of adaptive rationality suggests, incentive effects cannot be reliably assessed without controlling for the other parameters of the mode-selection threshold (accessibility, opportunity, cost etc.) and taking into account the interactive effects between variables. In particular, a high match potentially suppresses incentive effects and triggers the automatic mode; even when extrinsic motivation is very high. Depending on differences in accessibility, the experimental treatments are heterogeneous within one experimental condition. By measuring and accounting for the availability and accessibility of stored and relevant frames and scripts, this effect will be controlled in the present experiment.

stored mental schemata (a high activation weight) can, in the extreme case, completely suppress the effect of instrumental incentives. This is one plausible explanation for the inconsistent results of the high-stake studies cited above: none of the studies has explicitly taken into account interactions between accessibility and motivation. Thus, heterogeneous treatment effects within one experimental condition cannot be uncovered. Mean comparisons are inefficient because sub-groups react differently to one condition. Low accessibility subjects may be more prone to responding to incentive treatments than their high accessibility counterparts, who may not respond or even display opposite behavior. Presumably, this explains the reported increases in variance of the experimental data. Another caveat of most studies conducted so far is a relatively low number of observations. If incentive effects in experiments are of small effect size only, then a sufficient number of observations is necessary to reliably detect the treatment effects.

Concerning interpersonal trust, only a handful of experiments have explored the effect of motivation and stake-size in trust games or investment games. For example, Naef and Schupp (2009) do not find any significant reduction of amounts transferred in an investment game. Parco et al. (2002) find that increased stake size has a strong effect on trusting behavior in a 9-move centipede game and conclude that high monetary payments bring subject´s decisions closer to the equilibrium predictions of rational-choice models. Malhotra (2004) experimentally shows that a higher difference between the *status quo* and the "sucker" payoffs (an increase in the potential loss that the trustor incurs) negatively affects trust in a trust game. Johannson-Stenman et al. (2005) show that transfers in the investment game are significantly reduced in high-stake conditions, a finding also provided by Holm and Nysted (2008), who present evidence for a significant reduction of transfers in the trust game. They conclude that high stakes trigger an approximation to Nash-equilibrium because incentives "may induce the subjects to engage in more complex analysis instead of solely relying on their instinctive emotions concerning the choice at hand" (ibid. 532).[8]

In line with the experimental procedures developed by Johannson-Stenman (2005) and Holm and Nysted (2008), the *initial endowment* of the trustor and trustee will be varied on two levels. This affects the opportunity cost C_f of making a false decision in the automatic mode, and the potential loss that failed trust involves (see Malhotra 2004 for a related argument). If the

[8] Their explanation is surprisingly similar to an intuitive version of the model of adaptive rationality: According to Holm and Nysted, decisions in the trust game "can be thought of as solved on two different cognitive levels. On the first level [the automatic mode] the A-player decides if there is an alternative that has a direct attractiveness or makes the choice at random. At the second level [the rational mode], the player realizes the strategic situation, forms expectations about the other player and chooses the alternative that maximizes his utility function (which may not be entirely selfish). Those A-players reaching the second level may choose to trust or not, since they may be influenced by motivations (and expectations about motivations) such as inequality aversion, kindness, altruism reciprocity or efficiency. We believe that most A-players make their decision on the first level in the hypothetical treatment and on the second level when monetary motives are present" (2008: 532). Unfortunately, they do not further develop these ideas into a theoretical model.

trustee is in fact *not* trustworthy and distrust would have been the better option, the trustor always loses a higher absolute amount for a given relative level of trust. What is more, the gain and utility associated with a correct decision in the rational mode (U_{rc}) also increases because either correct decision (trust or distrust) can bring a higher absolute payoff in the high-stake condition, irrelevant of the trustee´s actual behavior. That is, both the *status quo* payoffs and the potential gain of a trustworthy response are higher than in the low stakes conditions. Therefore, the overall situational-extrinsic cognitive motivation to engage in a more rational and elaborate reasoning process should increase in the high-incentive condition.

In the experiment, the participants will receive either a high (40€) or a low (7€) initial endowment. The initial endowment of 7€ corresponds to the hourly wage of a student-assistant at the University of Mannheim (in the fall-semester of 2010). The high-stake condition introduces endowments which are 5.8 times higher than in the baseline condition. Given a duration of one hour per experimental session on average, a 40€ *status quo* payoff represents an hourly wage well above average. These high initial endowments aim to increase the likelihood of engaging in a more controlled elaboration of the potential risks involved in the choice of a trusting act *via* their effect on the motivation-component ($U_{rc}+C_f$) of the mode-selection threshold. Again, we can connect to the expected transfers X and Y using bridge hypotheses B1-B3. The transfers in the investment game and the observed levels of trust are expected to decrease in the high stake condition. We can derive auxiliary hypotheses concerning expectations of trustworthiness, the influence of instrumental variables and the expected decision times of the participants using the general model propositions P6-P8 (section 6.3). In particular, expectations should become more pessimistic; a controlled reasoning process warrants that expectations should approach the "equilibrium solution". Furthermore, with high initial endowments, the influence of instrumental variables should increase, and decision times should be relatively longer.

6.2.4. Participants

The experimental design was implemented using the z-Tree software package (Fischbacher 2007) for economic experiments. The experiment was conducted between August and November 2010 at the University of Mannheim. In a first wave, a sample of N=114 first-year sociology and political science students was recruited in class at the beginning of the fall semester. In the second wave, another N=184 students were recruited from the university´s experimental-subject pool and psychology classes. A dummy variable will be added to control for subject-pool equivalency (as suggested by Buchan et al. 2002).

Concerning the use of students as experimental subjects, note that a student sample is not representative (Naef & Schupp 2009). According to Levitt and List (2007), students generally act different in comparison to non-student groups. Harrison et al. (2007) find that students are

more risk averse and have less pronounced social preferences than non-students. Likewise, Gächter and Thöni (2004) show that students hold more pessimistic attitudes about trust, using the GSS survey items on generalized trust.[9] Bellamare and Kroeger (2007), using a representative sample in the Netherlands, show that students on average transfer *less* in an investment game, and also hold more pessimistic expectations of trustworthiness. In contrast, Naef and Schupp (2009), using a representative sample of German households (SOEP), show that students in fact send relatively *more* in an investment game. Thus, it is difficult to provide a definite answer to the question of potential systematic differences between a student sample and a representative sample with respect to trust. One advantage of a student sample is that the observations are relatively homogeneous on controls such as age or education.

There is trade-off between internal and external validity in the choice of method and design. Being able to conduct an experiment and resort to experimental data, in contrast to a survey study, provides high internal validity while compromising on external validity and representativeness of the data (Falk & Heckman 2009). But the strongest advantage of an experimental approach, that is, to test for causal effects and allow causal inference, far outweighs the relative disadvantages incurred with a limitedly representative dataset. On top of that, the experimental measures adopted in trust research in fact provide a considerable degree of external validity (Baran et al. 2010). Lastly, the choice of a student-sample in the experimental design is conforming to the current standard procedure of both experimental economic and psychological research.

6.2.5. Materials and Procedure

Upon arrival, the participants were seated in separate booths in the laboratory, where they would find a sheet of written general instructions about the procedures and the experiment (see Appendix B).

The instructions explained that they now participated in an experiment and could earn real money, the magnitude of which depended on their own decisions and the decisions of other participants. Participants were told that they had to perform several tasks, out of which *one* task was selected randomly at the end of the experiment to determine the final payoff. The money earned would be paid in cash directly after the experiment had ended. Participants were informed that communicating with others was strictly prohibited, and that the instructor could always help them if they silently raised their hands. The instructions made clear that at

[9] The three GSS items used are (1) "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"; (2) Do you think most people would try to take advantage of you if they got a chance or would they try to be fair"? and (3) "Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves"? On top of that, they ask for the frequency of behavioral manifestations of trust (i.e. lending a book, money, and leaving the door open) with similar results.

no point in time would the participants know the identity of others and *vice versa*, and that their decisions and payoffs remained completely anonymous.

Furthermore, the written instructions informed them that, at the beginning of the experiment, they would be randomly matched into a group (team) with another participant (partner). Thus, the context treatment was implemented from the very beginning of the experiment. While reading the instructions, participants could see the welcome screen on the computer in front of them, which presented the cooperative or neutral visual cues. With a click on the screen, participants could start working on the task.

On-screen instructions then presented the investment game scenario from the perspective of player A (the trustor), using the context treatment word-pairs consistently with the written instructions. On the next screen, participants were introduced to the decision interface and could make one "trial" decision for practice. The participants then had to answer ten control questions to make sure that they had understood how the payoffs of the game were determined.

Only when all ten questions were answered correctly could the participants proceed to the actual decision stage. The next screen asked for their transfer decision, followed by a screen to ask for their expectation of trustworthiness, that is, how much they expected to receive back from Player B.

When all participants had made their decision about the transfer in the role of Player A, they were informed about a re-start and presented with new on-screen instructions explaining the investment game in their new role as player B (the trustee). They were asked another four control questions to make sure that the new scenario was understood properly, and then had to make their reciprocal decision of trustworthiness, being presented with the transfer decision of a randomly selected participant. Lastly, they were asked their second-order expectations ("What do you think did player A expect to get back?").

Upon completion, several manipulation checks were collected. Subjects were asked three questions about the perceived importance of the decisions and whether much money was at stake, they were asked to rate the expected cooperativeness of the interaction partner and their expectations about the fairness of others, and they were asked a self-report about cognitive style. Then, participants had to answer the scales and survey items of generalized trust, the reciprocity norm, faith in intuition and need for cognition. All scales were elicited using a 7-point Likert-type scale (ranging from "fully agree" to "fully disagree").[10]

[10] Financial and logistic limitations prevented a course of action in which scale measures were collected *in advance* of the actual experiment In fact, asking survey items *at the end* of an experiment is a most common practice, both in psychological and economic experiments (a similar approach in trust experiments was taken by, for example, Buchan et al. 2002,

Finally, participants were asked all survey and control-variable items. They were then informed about the randomly selected task which would become payoff-relevant, which was either their choice as player A or B. The decisions by both participants were displayed and the final payoffs determined. The money was given to the participants in a separate room by calling out each computer-number and paying the participant in the other room.

6.3. Empirical Hypotheses

6.3.1. Using the Model to Predict Trust

In this section, we will derive testable empirical hypotheses to investigate the model of trust and adaptive rationality in an experimental setting. The hypotheses can be developed using the model propositions P1-P8, bridge hypotheses B1-B3, and the proposed effects of the experimental treatments. In addition to a number of direct main effects, the model allows to formulate more complex hypotheses about the interplay of several parameters and their interactions. In fact, one important benefit and advantage of the threshold model is that permits the specification of complex *interaction patterns*. Remember that the threshold for the unconditional choice of a trusting act was defined as:

$$m_i * a_j > 1 - C / (p * (U_{rc}+C_w))$$

To see how the threshold can be used to derive hypotheses about main effects and complex interaction patterns, the following assumptions (discussed in detail in the previous chapters) are collected into a set of bridge hypotheses to begin the analysis:

(1) $a_{j|i}=1$ (a script is temporarily accessible given the frame, A1)
(2) $a_{k|j}=1$ (an action is satisfactorily regulated given an activated script, A2)
(3) The automatic mode leads to a complete transfer of resources, X=E (B1)
(4) The rational mode supports any transfer between zero and the initial endowment, $X_{\varepsilon}[0, E]$, but transfers approximate the Nash-equilibrium of distrust X=0 (B2, B3)
(5) The decision times using the automatic mode are shorter than the decision times using the rational mode (P8.1, P8.2)
(6) $l_i = 1$ (the link between objects and chronically accessible frames is established)
(7) C<p*U (assumption to ensure that the threshold is well-behaved)

The most important element in this set of assumptions is the link between processing modes and transfer decisions, as stated in (3) and (4). As was argued before, the activation of the rational processing mode should lead, on average, to a decrease of observed levels of trust (see

Malhotra 2004 and Ermisch et al. 2009). While the implementation of a preliminary session in which to collect the survey measures before the actual experiment is methodically desirable, it is also more costly.

section 6.1.2). Using this set of bridge hypotheses, it is straightforward to derive empirical predictions regarding the effect and sign of the experimental treatments and accessibility measures that are elicited in the experimental design.

6.3.2. Main Effects

A number of general model propositions (P1-P8) can be directly translated into empirical hypotheses about statistical main effects. These main effects can be derived *ceteris paribus*, holding other parameters of the mode-selection threshold constant. For example, consider the impact of the chronic accessibility of frames and scripts on the definition of the situation and the subsequent choice of a trusting act. The model reveals that high accessibility increases the left-hand side (LHS) of the threshold, that is, the activation weight. The more readily available trust-related knowledge is to a trustor, the more likely it is that interpretation and choice in the trust problem occur in the automatic-mode:

H1 (frame accessibility): The higher is the chronic accessibility a_i of a trust frame, the higher is the probability of an unconditional choice of a trusting act in the automatic mode. With respect to transfer decisions in the investment game, this implies a positive main effect of frame accessibility. With respect to decision times, this implies a negative main effect.

H2 (script internalization): The higher is the chronic accessibility a_j of a trust-related script, the higher is the probability of an unconditional choice of a trusting act in the automatic mode. With respect to transfer decisions in the investment game, this implies a positive main effect of script internalization. With respect to decision times, this implies to a negative main effect.

The model of frame selection suggests that accessibility has an effect on the processing mode *via* its influence on the match and the "smoothness" of pattern-recognition. In the absence of more individuating information or social embeddedness, other categories of trust-related knowledge (such as specific expectations, reputation information etc.) are not appropriate and cannot be applied to the experimental situation. However, a relevant trust frame is provided by generalized interpersonal trust, an abstract relational schema that participants have or have not developed for interaction with anonymous others, and in the script of the reciprocity-norm. The focus on these two measures of trust-related knowledge has to be understood in relation to (and is motivated by) their interplay with the context treatment. The cooperative framing condition was designed to increase the appropriateness of a relational schema of generalized interpersonal trust and the validity of the norm of reciprocity. These knowledge structures can assist the solution of the trust problem in the anonymous one-shot experimental set-up, and their importance increases in the cooperative framing condition.

The model predicts main effects with respect to the two experimental treatments. Concerning the effect of the context treatment on the observed levels of trust, it is expected that the presence or absence of situational cues pointing towards the validity of a trust-related frame and script influence interpretation and choice in the trust problem.

H3 (context treatment): The cooperative context treatment increases the match m_i of a trust frame by providing relevant situational cues o_i. The higher is the situationally indicated appropriateness of a trust-related frame and script, the higher is the probability of an unconditional choice of a trusting act in the automatic mode. With respect to transfer decisions in the investment game, this implies a positive main effect of the cooperative framing condition. With respect to decision times, this implies a negative main effect.

The cooperative context provides cues that can be used by participants to interpret the situation more favorably than in the neutral framing condition. The word-pairs used in the cooperative framing condition´s instructions ("partner"/"team") are designed to decrease perceived social distance and signify that a communal relationship orientation and corresponding communal interaction norms (the reciprocity norm) are appropriate, whereas such cues are not presented in the neutral framing condition. The visual primes presented to the participants at the beginning of the experiment (a picture of hand-shakes, a picture of bank notes) supplement the word priming manipulation.

Next, the effect of the incentive treatment can be identified by looking at the mode-selection threshold. In particular, high initial endowments are expected to increase the cognitive motivation to engage the rational mode and push participants towards a rational processing of the trust problem during interpretation and choice. In particular, we can hypothesize that:

H4a (incentive treatment): High initial endowments increase cognitive motivation $U = (U_{rc}+C_f)$ to activate the rational mode. The higher the cognitive motivation U, the higher is the probability of a conditional choice of a trusting act in the rational mode. With respect to transfer decisions in the investment game, this implies a negative main effect of high initial endowments. With respect to decision times, this implies a positive main effect.

This hypothesis recasts a core postulate of dual-process theories, stating that motivation is a central determinant of the processing mode, and applies it to the present experimental set-up. Hypothesis H4 also comprises the "low-cost hypothesis" (Diekmann & Preisendörfer 1992, Rauhut & Krumpal 2008) of economic rational choice theory: attitude- and norm-conform behavior can only be expected in low-cost situations. The more is "at stake" for the actor, the higher is the likelihood that a rational processing of information will prevail in the stages of interpretation and choice. Then, social norms and attitude-conform behaviors are a mere part of instrumental cost-benefit considerations; they are not unconditional. Likewise, H4 implies

that unconditional trust can be disturbed whenever trust problems bear important consequences to the trustor. Consequentially, the use of heuristics such as choice-rules, subjective experiences and so forth should minimally occur in a controlled fashion, if their influence is not completely overridden by the intervention of the rational system.

It is important to note that the proposed main effect of instrumental incentives and high stakes conditions is mediated by a number of interactive effects which compensate it. As will be shown in section 6.3.3, the model posits that incentive effects can be fully suppressed if, for example, actors´ chronic accessibility of relevant frames and scripts is high. In other words, even in high-cost situations, actors can be "immune" against instrumental incentives and stick to the mental schemata and heuristics suggested by the associative memory-system; even when the stakes are very high and cognitive motivation prompts towards a more rational elaboration of the selection problem. A corollary hypothesis can be formulated with respect to the individual-intrinsic dimension of cognitive motivation, which will be measured using the NFC/FI scales.

H4b (intrinsic motivation): The higher the intrinsic cognitive motivation U, the higher is the probability of a conditional choice of a trusting act in the rational mode. With respect to transfer decisions in the investment game, this implies a negative main effect of intrinsic motivation. With respect to decision times, this implies a positive main effect.

One potential source of variance in intrinsic cognitive motivation is the "Faith in Intuition" and "Need for Cognition" of the participants, which will be elicited with the corresponding scales in the survey-stage of the experiment. As was argued in section 6.1.5, a number of studies have provided evidence that processing preferences accumulate into different cognitive types, based on NFC/FI classifications. Traditionally, NFC is interpreted to measure a stable individual-intrinsic aspect of motivation. Thus, it is possible to predict individual-intrinsic differences in the way subjects chronically activate a certain processing mode, and control for a "preference" for processing, by keeping track of the differences in NFC/FI. In short, it is hypothesized that unconditional trust is more common among low motivation-type individuals than among high motivation-type individuals.

Hypotheses H1-H4 can be tested using the present design. Next, all hypotheses which cannot (or can only partially) be tested with the present experimental design will be derived. They are stated here for the sake of completeness. With respect to the cost and effort C associated with elaborate interpretation and choice in the trust problem, the model implies:

H5 (effort): The higher is the cost and mental effort C associated with rational processing, the higher is the probability of an unconditional choice of a trusting act. With respect to transfer

decision in the investment game, this implies a positive main effect of cost and mental effort C. With respect to the decision times, this implies a negative main effect.

Since task complexity does not vary between treatments, mental effort will mainly differ on its individual-intrinsic dimension. However, a direct measure of mental effort, such as cardio-vascular response and neural activity (see Fairclough & Mulder 2011) was not scheduled. Therefore, H5 will not be analyzed with the experimental data. Considering opportunities p for engagement of the rational mode, the model implies:

H6 (opportunity): The lower is available opportunity p necessary for rational processing, the higher is the probability of an unconditional choice of a trusting act in the automatic mode. With respect to transfer decision in the investment game, this implies a negative main effect of opportunity p.

That is to say, restrictions on the scarce resource of attention and cognitive capacity, or direct time-pressure (all impact opportunity p negatively) prevent the activation of the rational mode. Then, the trustor has to rely on heuristic shortcuts to make a decision about trust. For example, in the lack of further individuating information, the trustor can resort to cognitive or affective feelings, use simplifying choice rules (i.e. a coin toss) or application of heuristic schemata ("doctors can always be trusted") when there is no opportunity. In contrast, the activation of the rational mode is feasible when opportunity p is sufficient. It opens up the potential for the development of conditional trust. A hypothesis about the effect of opportunity on decision times was omitted here. This is because opportunity, in the form of time-pressure, limits itself the amount of available decision time. As with effort C, the effect of opportunity will not be analyzed in the current set-up: subjects were given an unlimited amount of time to think at every stage of the experiment, and at no point in time capacity was limited with concurrent task-activities.

6.3.3. Interaction Effects

To analyze interaction effects between model parameters, it is necessary to vary more than one parameter at a time and track the effect of a parameter change on the threshold value, while introducing some variation in other parameter values (see Kroneberg 2011b for further discussion). In what follows, the focus is on those interaction effects that can be tested and quantified using the experimental design. This course of action is exemplary; all other higher-order interactions can be derived analogously. In particular, the analysis focuses on the effect

of the experimental treatments on the threshold value and their impact on the total balance of the mode-selection threshold while simultaneously varying chronic accessibility.[11]

The experimental treatments change two parameters of the threshold. First, the cooperative *versus* neutral context is designed to influence the presence of situational cues $o_{i,}$ as part of the match $m_i = m_i(o_i)$. Second, the high *versus* low incentive treatment is designed to manipulate cognitive motivation $U = (U_{rc}+C_w)$. What does the model tell us about the interaction between the two parameters, the interaction between each parameter and the chronic accessibility a_j of a reciprocity script, and the joint interplay of all three variables? Neglecting all constant parameters for the moment, we can write:

$$\mathbf{o_i} * \mathbf{a_j} > 1 - S / \mathbf{U}$$

S is the constant derived from (C/p). Obviously, the threshold depends on all three parameters at the same time, and whether a single parameter change "tips over" the threshold balance crucially depends on the specification of all other parameter values. That is to say, the model predicts two- and three-way interactions between U, o_i and a_j. In a statistical model, we would have to include not only main effects U, o_i, and a_j, but also interaction terms ($U*o_i$), ($U*a_j$), ($a_j*o_{i'}$) and the three-way interaction ($U*a_j*o_i$). But what is the predicted sign of these effects?

As presented in full detail in Appendix C, the model can be used to predict the statistical sign of all interaction effects and their direction with respect to the expected transfers in the investment game using bridge hypotheses B1-B3. The set of valid combinations that remain includes 17 different outcome patterns. The following table summarizes all the predicted *interaction patterns*, including the sign of the main-effects, second- and third-order interactions with respect to the transfer decision in an investment game. These patterns are consistent with the model of adaptive rationality, specifying the interaction between script accessibility a_j, situational cues o_i and cognitive motivation U (table 4):

[11] Since all accessibility parameters are located on the left-hand side of the mode-selection threshold (directly influencing the activation weight), their effect and interactions with other threshold parameters are identical. Therefore, the following hypotheses, related to the chronic accessibility a_j of a script, can be easily restated in terms of chronic frame accessibility. The predicted sign and effect of interactions with other parameters is the same.

Table 4: Predicted interaction patterns for *reltrust*

	Predicted Interaction Patterns (Main- and Interaction Effects)																
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a_j	0	≥ 0	≥ 0	≥ 0	$= 0$	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	0
U	0	≤ 0	≤ 0	≤ 0	< 0	$= 0$	$= 0$	≤ 0	≤ 0	≤ 0	< 0	$= 0$	≤ 0	≤ 0	$= 0$	≤ 0	0
o_i	0	≥ 0	≥ 0	≥ 0	$= 0$	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	$= 0$	> 0	≥ 0	≥ 0	≥ 0	≥ 0	0
$U \cdot o_i$	0	≥ 0	> 0	≥ 0	$= 0$	$= 0$	≥ 0	≥ 0	≥ 0	≥ 0	$= 0$	$= 0$	≤ 0	< 0	$= 0$	≤ 0	0
$a_i \cdot U$	0	≥ 0	$= 0$	≥ 0	$= 0$	$= 0$	≤ 0	≤ 0	≤ 0	≤ 0	$= 0$	$= 0$	≥ 0	$= 0$	$= 0$	≤ 0	0
$a_i \cdot o_i$	0	≤ 0	$= 0$	≥ 0	$= 0$	< 0	≤ 0	≤ 0	≤ 0	≥ 0	$= 0$	$= 0$	≥ 0	$= 0$	> 0	≥ 0	0
$a_i \cdot o_i \cdot U$	0	≤ 0	$= 0$	> 0	$= 0$	$= 0$	> 0	> 0	> 0	> 0	$= 0$	$= 0$	> 0	$= 0$	$= 0$	< 0	0

Note: The table presents predicted interaction patterns between chronic script accessibility a_j, situational cues o_i and motivation U to predict transfer decisions in the investment game.

The model admits a number of different interaction patterns between the model parameters. Depending on the parameter values and the effect of the treatments on the threshold, the expected higher-order interactions can be zero, positive or negative. In contrast, the sign of the main effects is unambiguous. The three-way interaction can be positive *or* negative, depending on the joint effect of the parameters (see Kroneberg 2011b for a detailed discussion). A positive three-way interaction results whenever a high value of all parameters determining the left-hand side (the activation weight) is *necessary* to trigger the automatic mode and counterbalance the negative effect of cognitive motivation U. In this case, a cooperative context *and* high script accessibility a_j are necessary to reduce the negative incentive effect pushing towards the rational mode. A negative sign of the three-way interaction indicates that one of the two components is already *sufficient.* That is, when facing high incentives and motivation to engage the rational mode, actors choose unconditional trust *either* when the context is cooperative *or* when accessibility is high.

The result of this analysis is confusing at first glance. Depending on the concrete parameter values and treatment effects, the predicted interaction patterns are considerably diverse. This is not to say, however, that the model admits and predicts statistical interaction effects at random and without any restrictions. Even when the falsification of one particular interaction effect, detached and separated from the joint set of other hypotheses in the interaction pattern, is factually not possible (i.e. the model admits both a positive, a negative and a zero three-way interaction), the overall interaction patterns which can be predicted provide a set of admissible data patterns, against which any empirical deviation can be regarded as negative evidence of the theoretical model.

When analyzing the data, it is important to keep in mind that the data represent an estimated aggregate effect from a distribution of individual threshold values, and the statistical results mirror the "average" parameter constellations found in the sample. This means that the data

may not reveal an overall consistent pattern if the data is too heterogeneous with respect to the distribution of individual threshold values (Kroneberg 2011b). Using a relatively heterogeneous student population (i.e. a sampling of primarily first-year students) thus could engender a relative methodological advantage, because the heterogeneity in threshold-values is presumably smaller than in a population-representative sample. Let us now take a closer look at the resulting interaction effects. In particular, the following two- and three-way interaction hypotheses are implied by the model, holding other parameters constant:

*H7a (H2 x H4): The effect of situational cues (cooperative versus neutral framing condition) is mediated by the chronic accessibility a_j of a trust-related script. Both parameters are necessary for a high activation weight. The higher is the chronic accessibility of the script, the stronger is the effect of the context treatment. With respect to transfer decisions in the investment game, this corresponds to a positive two-way interaction (oi * a_j). With respect to decision times, this corresponds to a negative two-way interaction.*

This formulation of the two-way interaction does not account for the third variable. As can be seen from the predicted interaction patterns, the two-way interaction (o_i * a_i) can well be negative or zero once the incentive variable is accounted for. However, when varying the two components of the match only, the above formulation is accurate. We can make a similar prediction with respect to the chronic accessibility a_i of a trust frame:

*H7b (H1 x H4): The effect of situational cues o_i (cooperative versus neutral framing condition) is mediated by the chronic accessibility a_i of the generalized trust frame. Both parameters are necessary for a high match. The higher is the chronic accessibility of the generalized trust frame, the stronger is the effect of the context treatment. With respect to transfer decisions in the investment game, this implies a positive two-way interaction (oi * ai). With respect to decision times, this implies a negative two-way interaction.*

Hypotheses H7a and H7b do not take into account cognitive motivation U. The interaction patterns presented in table 5 predict positive *and* negative two-way interactions between cues o_i and accessibility a_j. The hypothesis of a positive interaction between cues and accessibility, as stated above, must be qualified when varying more than two interacting components of the threshold simultaneously. In sum, the model predicts that the effect of the symbolical cues presented in the context depends on the chronic accessibility of appropriate knowledge structures to "decode" them. Thus, the context treatments are expected to have no effect for those subjects who report low chronic accessibility of trust-related frames and scripts and therefore cannot make use of the cues presented during interpretation. This hypothesis contrasts to generic dual-process models (Mayerl 2010). While generic dual-process models predict an *in*creasing effect of situational cues with *de*creasing accessibility, the model of frame selection posits that knowledge must first and foremost be latently accessible to allow for correct inter-

pretation of the corresponding cues. One plausible explanation for this discrepancy in theoretical predictions is that dual-process accounts have traditionally focused more on the impact of temporary accessibility (as demonstrated, for example, in the priming research-paradigm), while the model of frame-selection emphasizes the importance of chronic accessibility, that is, the "strength" of internalization and the latent activation-potential of knowledge.

Next, the model predicts an interaction between chronic accessibility a_j and cognitive motivation U. This hypothesis is a particular interesting one because it contrasts to standard economic models:

*H8 (H1 x H4): The effect of cognitive motivation U (H4) is mediated by chronic script accessibility a_j. High chronic accessibility increases the activation weight of the mode-selection threshold. The higher is the chronic accessibility a_j of a trust-related script (i.e. the reciprocity norm), the weaker is the negative effect of cognitive motivation U. With respect to transfer decisions in the investment game, this implies a positive two-way interaction (oi * ai). With respect to decision times, this implies a negative two-way interaction.*

In other words, the negative effect of high initial endowments (an increase in cognitive motivation U, pushing subjects towards rational elaboration and conditional trust), can be counterbalanced and even fully suppressed if subjects have internalized a regulative script which can be applied to the current situation. High chronic accessibility promotes high activation weights, and therefore an automatic application of stored knowledge without further consideration of instrumental factors. If the automatic mode prevails during mode-selection as a result of a high match, then the decision-making process does not follow the principles of economic utility maximization anymore. Actors chose actions solely based on the selection rules of the automatic mode (selecting frames, scripts and actions with the highest activation weight). The "logic of appropriateness" then unfolds in the patterns of spreading activation that the perception of situational stimuli affords. This also implies that:

*H9 (H3 x H4): The effect of cognitive motivation U (H4) is mediated by the presence of situational cues o_i. A cooperative framing of the trust problem increases the activation weight of trust-related frames and scripts. The higher is the appropriateness of trust-related knowledge, as indicated by cues o_i, the weaker is the negative effect of cognitive motivation U. With respect to transfer decisions in the investment game, this implies a positive two-way interaction (oi * ai). With respect to decision times, this implies a negative two-way interaction.*

Principally, this means that the experimental treatments cannot be analyzed independently. Since situational cues o_i and accessibility parameters a_i, a_j and $a_{j|i}$ work in the same direction with respect to their influence on the activation weight and mode-selection threshold, the predicted sign of the two-way interaction between cues and motivation is identical to H6. High

initial endowments, *via* their effect on cognitive motivation U, increase the probability of an activation of the rational mode and, in consequence, promote conditional trust. However, these incentive effects can be diminished by the presence of situational cues indicating the appropriateness of stored schemata because they influence activation weights. If the initial categorizations are supported by a detection of relevant external cues, then appropriateness remains unquestioned and information processing occurs at the default automatic mode. Lastly, the model predicts a three-way interaction between cues o_i, motivation U, and accessibility a_j:

H10 (H1 xH3 x H4): The negative effect of cognitive motivation U (H4) is jointly mediated by chronic accessibility a_j of a trust-related script and the presence of situational cues o_i. This implies a three-way interaction, the sign of which depends on the joint effect of a_j and o_i on the threshold. The three-way interaction will be negative if a high value of either parameter is sufficient to counterbalance the effect of motivation on the threshold. It will be positive if a high value of both parameters is necessary to compensate the negative effect of motivation

Once the three-way interaction is statistically taken into account, the predicted two-way interactions will change according to the predicted interaction patterns, as presented in table 5 above. Hypothesis H10 completes the set of empirical predictions that the model affords with respect to the observable outcomes in the investment game context and when varying accessibility, situational cues and motivation at the same time. It is important to keep in mind that the analysis of interaction effects is a model-inherent requirement and necessity, motivated by the model of trust and adaptive rationality. They are not conducted *ad libitum*. The theoretical framework developed here suggests that an analysis of simple main effects will most likely *not* be sufficient to properly describe the data, because interactions between the processing mode determinants are an ever present facet of interpretation and choice.

6.4. Descriptive Statistics

A total of N=298 participants were recruited and participated in the main experiment between August and November 2010. Participants were recruited in class at the beginning of the fall-semester and randomly selected into the different experimental conditions. The experiment was conducted in 24 separate sessions, which lasted about one hour each. The average group size was about 12 participants per session, and participants earned an average of 18€ from their participation in the experiment (table 5):

Table 5: Experimental conditions and number of observations

Treatment s		Context		Total
		Neutral	Cooperative	
Incentives	Low (7€)	76	70	146
	High (40€)	76	76	152
	Total	152	146	298

The number of observations across cells and experimental conditions is almost balanced. A lower number of observations in the "Low/Cooperative"-condition resulted due to random fluctuations. The next table summarizes basic information of the main dependent and independent variables used in the empirical analysis (table 6):

Table 6: Summary statistics of dependent and independent variables

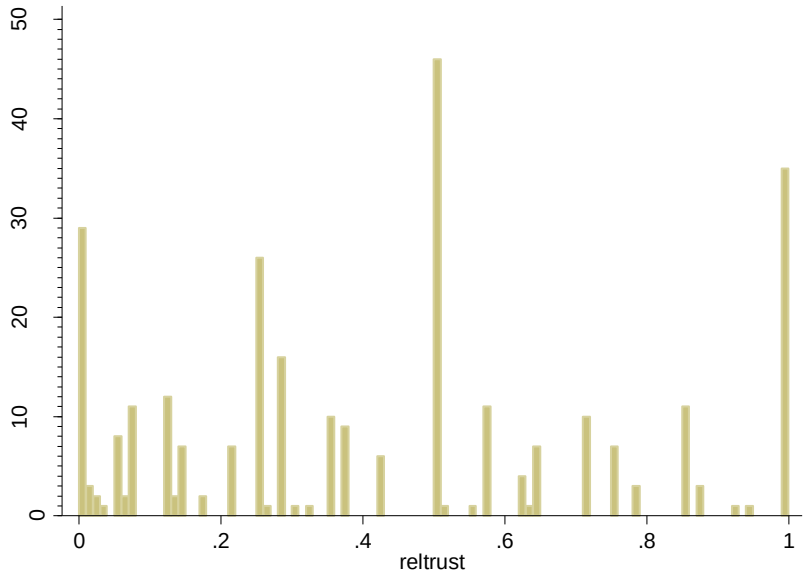
Label	Measure / Operationalization	[Min, Max]	Mean	Std. Dev.
Dependent Variables				
reltrust	Relative amount sent in investment game, X/E	[0, 1]	0.43	0.31
time	Decision time in seconds	[2.95, 207]	17.95	17.84
logtime	Logarithm of decision time	[1.08 5.34]	2.64	0.66
Independent Variables				
end	Incentive treatment (0=low, 1=high, dummy)	[0, 1]	-	-
frame	Context treatment (0=neutral, 1=coop., dummy)	[0, 1]	-	-
trustscale	Interpersonal trust (Kassebaum 2004)	[0.16, 0.89]	0.57	0.13
recscale	Norm of reciprocity (Perugini et al. 2003)	[0.36, 0.91]	0.68	0.09
fiscale	Faith in intuition (Keller et al. 2000)	[0.24 ,0.93]	0.65	0.13
nfcscale	Need for cognition (Keller et al. 2000)	[0.41, 0.97]	0.77	0.12
age	Respondent age in years	[18, 43]	21.89	3.77
sex	Respondent gender (0= male, 1= female, dummy)	[0, 1]	0.57	0.49
partner	Relationship status (0=no 1= yes, dummy)	[0, 1]	0.48	0.50
partnerl	Relationship length in months	[0, 200]	25.05	23.41
income	Income response categories	[0, 1875]	535.93	288.87
semester	Respondent semester´s studied	[0, 20]	3.36	3.66
append	Recruitment wave (0=first, 1=second, dummy)	[0, 1]	-	-

Concerning the composition of the sample, a total of 157 participants (52%) were first-semester students, another 65 participants (22%) had completed their third semester. The largest groups within the sample comprised sociology (18%) and political science (18%) degrees followed by business sciences (13%), economics (8%), psychology (8%) and IT sciences (8%). The participants were 21.9 years of age on average (SD=3.77), and a little more than half (57%) of the participants were female. About half of all (48%) reported to be currently engaged in a relationship. In this group, relationships had been continuing for about 25

months (SD=23.4). Participants reported an average monthly gross income of 535€ (SD=288€), including a maximum reported income of 1875€. Looking at social background, 59% of the participants indicated that their father had completed upper secondary school ("Abitur"), while 15% each reported their father´s educational background as middle-school ("Realschule") and lower secondary school ("Hauptschule"). These categories are the main educational degrees of the German educational system (see Müller et al. 1998). The average reported student income significantly differs between high and low social-backgrounds. Students whose father had completed upper secondary school report an average income of 573€ (SD=289€), which drops to an average 456€ ("Realschule", SD=218€) and 464€ ("Hauptschule", SD=302€), respectively.

Across treatments and conditions, trustors transferred an average of 43% of their initial endowment to the trustee. Transfers spanned the whole range from zero to full transfers. That is, both complete trust and distrust can be observed in the sample. About 10% (N=29) of the participants opted for the safe alternative of distrust and transferred none of their initial endowment. Another 12% (N=35) transferred the full initial endowment. The following graph depicts a frequency histogram of the dependent variable *reltrust* (figure 21):

Figure 21: Frequency histogram of *reltrust*

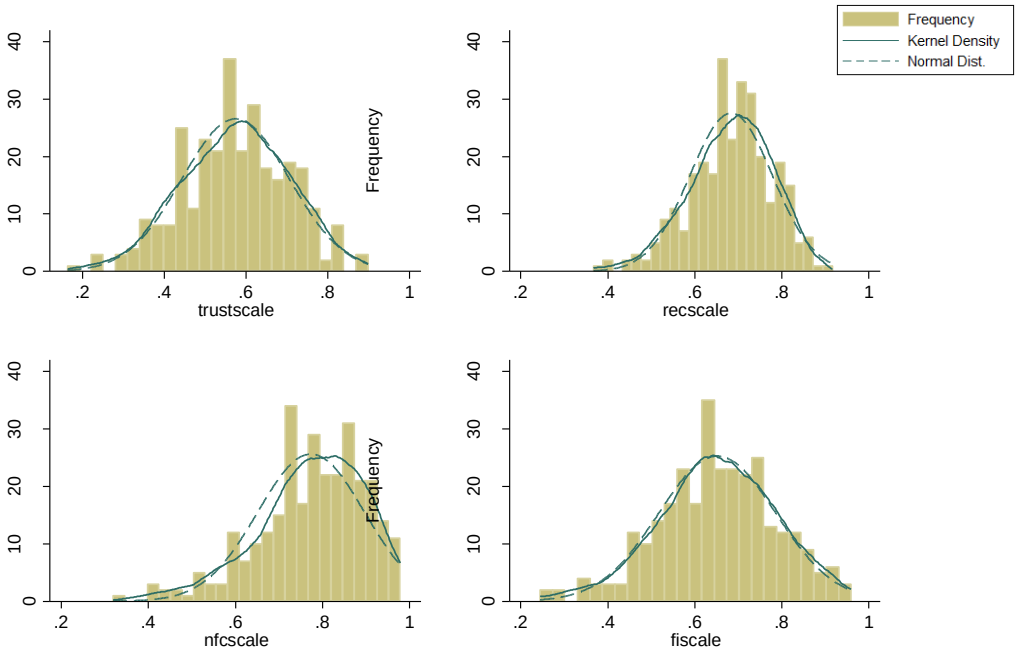


As can be seen from the histogram, the mode of the distribution is at a relative transfer of half of the endowment, which about 15% (N=46) of the participants opted for. Overall, about 52% (N=157) of the trustors transferred less than half of their endowment, while about 33% (N=95) transferred more than half of their endowment. Looking at the distribution of *reltrust*, it is immediately apparent that the variable is not normally distributed (Shapiro-Wilk normality test, Z=4.53, p<0.001). With respect to statistical inference, this warrants some caution when using parametric tests that rely on normality assumptions, such as estimation of confi-

dence intervals and t-tests, and it encourages the use of statistical models that can handle "heavy tails" at the fringes of the distribution.

Concerning the distribution of the four continuous independent variables, the following graph combines histograms of the measures of (1) chronic accessibility of a trust-related frame (2) chronic accessibility of a trust-related script, (3) faith in intuition and (4) need for cognition. A kernel density estimate and a normal density plot were added to each graph (figure 22):

Figure 22: Frequency histogram of (1) *trustscale*, (2) *recscale*, (3) *nfcscale*, (4) *fiscale*



All scales were normalized to the unit interval. The distributions of the chronic accessibility and NFC/FI measures appear to be normal from the graphs. In fact, a Shapiro-Wilk normality test cannot reject the null-hypothesis of a normal distribution for *trustscale* (Z=-0.85, p=0.8) and *fiscale* (Z=0.5, p=0.3). However, it does so for *nfcscale* (Z=5.1, p<0.001). The distribution´s mean is .77 with a standard deviation of .12; the distribution is negatively skewed and left-tailed. A closer inspection of the underlying NFC-scale items reveals that response frequencies are skewed towards the "high"-end (strong agreement with high intrinsic cognitive motivation) of the scale. There are two explanations for this finding: (1) it might be that the student sample population generally has a high intrinsic cognitive motivation, or (2) the used scale items cannot properly discriminate between low and high NFC-individuals of the sample. In either case, the result is a loss of discriminative power and information due to a potential and non-linear "ceiling" effect, which has to be respected. Furthermore, the distribution of the norm of reciprocity scale looks well-behaved from the histogram, but the null-hypothesis of a normal distribution of *recscale* can be rejected (Shapiro-Wilk normality test, Z=2.1, p=0.02). A skewness/kurtosis test (D´Agostino et al. 1990) reveals that *recscale* is negatively

skewed, while its kurtosis is not different from normal ($\chi^2(2, N=298) = 8.35$, p=0.015). However, note that all four variables will be used as independent variables in the subsequent analysis; this relaxes most of the distributional concerns with respect to statistical inference.

In order to get a first impression of the relation between dependent and independent variables, the next table displays the conditional means of *reltrust*, as well as test statistics for two-sided, non-parametric Wilcoxon rank-sum tests ("WRT"). For each descriptive statistic (as well as the estimates from regression models below), statistical significance is specified at the 5%-level for two-sided tests unless stated otherwise. To create conditional means for the continuous measures of chronic accessibility and FI/NFC, the sample was split along the median of the corresponding response variable to create high- and low-score groups (table 7):[12]

Table 7: Conditional mean of *reltrust* within subgroups

Variable	Conditional mean of *reltrust* (standard deviation)		Wilcoxon rank-sum test	
			Z=	*p*=
end	Low 0.49 (0.32)	High 0.38 (0.31)	2.85	0.004
frame	Neutral 0.43 (0.29)	Cooperative 0.43 (0.34)	0.47	0.63
trustscale	Low Trust 0.40 (0.32)	High Trust 0.46 (0.32)	-1.81	0.071
recscale	Low Reciprocity 0.41 (0.30)	High Reciprocity 0.45 (0.34)	-1.11	0.27
fiscale	Low FI 0.44 (0.33)	High FI 0.42 (0.30)	0.52	0.74
nfcscale	Low NFC 0.41 (0.33)	High NFC 0.45 (0.31)	-1.34	0.17
sex	Male=0 0.47 (0.36)	Female=1 0.41 (0.28)	1.09	0.27
partner	No Partner 0.43 (0.32)	Partner 0.43 (0.31)	-0.04	0.97

Using the information in the table, we can assess simple main effects and conduct a preliminary examination of hypotheses H1-H4.[13] For instance, according to hypothesis H1, the probability of unconditional trust should increase with chronic frame accessibility. Using *trust-*

[12] While the experimental treatment variables *end* and *frame*, as well as controls *sex* and *partner*, are binary by nature, this is not the case for *trustscale, recscale, nfcscale* and *fiscale.* It is now widely accepted that dichotomization of continuous variables can introduce a number of unwanted side-effects, such as loss of effect size and statistical power or introduction of potential artifacts (Cohen 1983, McCallum et al. 2002). For presentational purposes, a mean-split is conducted. The continuous measures will be analyzed with regression techniques in the next section.

[13] To prevent accumulation of the family-wise type-1-error, one can adjust α-levels using the Bonferroni procedure. Given that N=8 mean comparisons were statistically computed, the appropriate p-value is 0.05/8=0.00625 for a significance level of α=.05. A quick look at table 8 reveals that only *end* has a significant effect on *reltrust* when using Bonferroni-adjusted significance levels; the results do not change for any variable.

scale as an indicator to form subgroups across treatments, the conditional mean of *reltrust* increases from M=0.40 to M=0.46 between low and high chronic frame accessibility. However, this difference is not significant (WRT, Z =-1.81, p=0.071). That is, when tested separately and across conditions, a comparison of the conditional mean of *reltrust* does not support H1. Likewise, *reltrust* does not differ between high- and low chronic script accessibility, using *recscale* as the grouping variable (WRT, Z=-1.11, p=0.27). Thus hypothesis H2, postulating a main effect of script internalization, cannot be supported by this preliminary test either. But while there is no measurable effect in *reltrust* across high- and low-reciprocity groups, the variance in the data slightly increases in high-reciprocity subjects, the increase being marginally significant (Levene´s robust test, F(1, 296) = 3.57, p=0.059). Thus, high-reciprocity subjects may be more heterogeneous in their response to the experimental treatments than low-reciprocity subjects, a finding which points to a potential interaction between chronic script accessibility and other variables, such as the initial endowment and framing-conditions.

With respect to the experimental treatments, the conditional distribution of *reltrust* does not differ between framing conditions. In both cases, trustors transferred about 43% of their initial endowment (WRT, Z=0.47, p=0.63). Thus, the comparison of conditional means does not support main effect hypothesis H3, stating that a cooperative context increases unconditional trust across all conditions. Again, there is an interesting twist to this result: the variance in the data significantly increases in the cooperative framing condition (Levene´s robust test, F(1, 296) = 5.44, p=0.02). While no main effect can be observed at first glance, the result indicates that the observations are heterogeneous in their response to the framing condition. As suggested by the model of trust and adaptive rationality, the effect of the framing-treatment may depend on other parameters of the mode-selection threshold. This warrants a consideration of potential moderators such as chronic accessibility or cognitive motivation.

On the other hand, the comparison of conditional means reveals that *reltrust* is significantly lower with high initial endowments. Across framing conditions, the relative transfers decrease from an average of M=0.49 to an average of M=0.38 when the "stakes are high" (WRT, Z=2.85, p=0.004). This observation is in line with hypothesis H4a, stating that high initial endowments foster conditional trust. Obviously, whether or not the trust problem includes high or low "stake sizes" does matter to trustors. This observation suggests that there is strong and direct effect of the incentive structure of a trust problem on cognitive motivation. At the same time, intrinsic cognitive motivation does not have the same negative effect: in the high-NFC group, the conditional mean of *reltrust* slightly *in*creases from M=0.41 to M=0.45 across all conditions, but the hypothesis that the underlying distributions are the same cannot be rejected (WRT, Z=-1.34, p=0.17). Thus, with respect to hypothesis H4b, the preliminary analysis does not yield a conclusive result. Looking at the "intuitive" counterpart of the NFC-scale, there is no noticeable difference between high- and low-FI subjects (WRT, Z=0.52, p=0.74).

The following table summarizes average transfer decisions of the trustors separated by experimental conditions. As indicated by the analysis above, the conditional mean of *reltrust* significantly differs between high and low endowment conditions when holding *frame* constant, but statistically significant effects cannot be observed between a neutral and cooperative framing when holding *end* constant. Notably, average transfers drop to M=0.37 in the high/cooperative-condition, revealing the lowest average level of trust in the sample (table 8):

Table 8: Conditional mean of *reltrust* within experimental treatment groups

Treatment / Level		Context	
		Neutral	Cooperative
Incentives	Low (7€)	0.47 (0.29)	0.50 (0.35)
	High (40€)	0.40 (0.31)	0.37 (0.32)

Note: Table presents means of *reltrust* conditional on treatment factors. Standard deviations in brackets.

The data also point towards a common result in experimental trust research: a difference between male and female participants in the levels of trust. The conditional mean of *reltrust* drops from M=0.47 for males to M=0.41 for females (see table 7). However, this difference is not significant across conditions (WRT, Z=1.1, p=0.27). At the same time, the relative transfers of male participants show significantly more variation than the responses of females (Levene´s robust test, F(1, 296)=14.91, p<0.001). Overall, *some* effect of gender seems to be present in the data, weakly confirming other results, but the sample presents merely a congruent "tendency" and the tests do not corroborate gender effects with sufficient certainty.

Concerning the effect of an ongoing partnership on the level of *reltrust*, the data do not provide any evidence that there is a difference between single and non-single subjects (WRT, Z=-0.04, p=0.97). The inclusion of the variable *partner* was initially based on the hypothesis that relationship status affects the chronic accessibility of trust-related frames and scripts. The effect of relationship status may transpire only indirectly *via* its influence on the *trustscale* and *recscale* variables. In fact, there is a marginally significant difference between single (M=0.59, SD=0.12) and non-single (M=0.56, SD=0.14) subjects in the measure of trustscale (two-sided t-test, t(296)=1.77, p=0.077).[14] Surprisingly, chronic frame accessibility is *lower* for subjects being currently engaged in a relationship. It is hard to assess the substantial meaning of this result. As speculated by Ermisch et al. (2009), single subjects might have a greater incentive for interaction with strangers and therefore develop an overall more positive attitude of generalized interpersonal trust. However, this does not translate into an overall main effect on *reltrust*. On the other hand, there is no effect of being in a relationship on chronic script accessibility. The conditional means of *recscale* do not differ between low reciprocity

[14] The t-test was used in this instance because *trustscale* is sufficiently normally distributed.

(M=0.69, SD=0.09) and high reciprocity subgroups (M=0.67, SD=0.09; two-sided t-test, t(296)=1.25, p=0.21). Overall, the effect of *partner* is negligible.

6.5. Analyzing Trust

6.5.1. Model Specification

The preceding descriptive analysis of the experimental data is well-suited to assess the basic tendencies and major characteristics of the sample. But its informative value is limited in testing the model of trust and adaptive rationality. For one, it is necessary to pay attention to interactive effects between threshold-parameters when analyzing the data. This is difficult to accomplish using discrete mean comparisons and one-parameter tests. Most importantly, the separate testing of higher-order interactions and main effects using discrete tests drastically increases the number of necessary tests to be performed, increasing the overall type-1-error probability. A multiple test is advisable. Second, the use of dichotomizations for continuous variables is rarely justified. In fact, a common side-effect of transforming continuous variables into binary dichotomies for statistical inference is a loss of information, statistical power, and the potential introduction of artifacts; it is inferior to the use of continuous quantitative data which are preferable whenever within reach (Cohen 1983, McCallum 2002).

Therefore, the following analyses will make use of multiple regression techniques to analyze the joint effect of independent variables on *reltrust*. In order to estimate and test the interaction patterns which were derived in chapter 6.3.3, the mode-selection threshold parameters will be fed into a linear regression model that explains the expected value of *reltrust*, the relative transfer of the trustor, as a function of experimental treatment conditions, chronic accessibility measures and their interactions, holding constant any other parameters for which a measure was elicited. Note that, even with the present data, the analysis can only cover a partial test of the model of trust and adaptive rationality. The most obvious reason for this limitation is that, as the number of varying parameters increases, the interaction patterns (1) become increasingly complex and (2) their analysis requires a large sample size in order to provide a sufficient amount of observations across cells. In section 6.3.3, interaction patterns have been derived for $a_{j,}$ the chronic accessibility of a trust-related script, situational cues o_i and cognitive motivation U. Using these parameters, a linear model can be specified as:

(1) $E(reltrust|\mathbf{x}) = \mathbf{x\beta} + e = \beta_0 + \beta_1 * end + \beta_2 * frame + \beta_3 * recscale + \beta_4 * end * frame + \beta_5 * frame * recscale + \beta_6 * end * recscale + \beta_7 * end * frame * recscale + trustscale + \text{controls} + e$

All independent variables were discussed and introduced above. The variable *trustscale* is introduced into the model to hold constant its influence on the activation weight while analyzing

the remaining parameters. Moreover, the measures of "faith in intuition" and "need for cognition" will be added to the set of control variables, as discussed in section 6.1.5.

Using chronic accessibility a_i of a trust-related frame instead to specify a second model, the predicted interaction patterns do not change, since the effects of a_i and a_j on the value of the mode-selection threshold are identical. Thus a second model can be specified by interchanging *trustscale* and *recscale* in all of the above terms:

(2) $E(reltrust|\mathbf{x}) = \mathbf{x\beta}+e = \beta_0 + \beta_1 * end + \beta_2 * frame + \beta_3 * trustscale + \beta_4 * end * frame + \beta_5 * frame * trustscale + \beta_6 * end * trustscale + \beta_7 * end * frame * trustscale + recscale$ + controls + e

This analysis can also be adapted to analyze the joint effect of the parameters which define the match, holding the effect of cognitive motivation constant. In other words, both *trustscale* and *recscale* will be varied across framing conditions. In this case, a third linear model can be specified as:

(3) $E(reltrust|\mathbf{x}) = \mathbf{x\beta} + e = \beta_0 + \beta_1 * frame + \beta_2 * trustscale + \beta_3 * recscale + \beta_4 * trustscale * frame + \beta_5 * frame * recscale + \beta_6 * trustscale * recscale + \beta_7 * trustscale * frame * recscale + end$ + controls + e

Hypotheses for this model have not been analytically derived here, but this can be done similar to the procedure discussed in Appendix C. Principally, note that all three parameters which vary in the third model specification (chronic frame accessibility, chronic script accessibility, and situational cues o_i) are located on the left-hand side of the mode-selection threshold and jointly define the activation weight. While each of the parameters is predicted to have a positive simple main effect, the joint interaction patterns and the direction of the predicted sign of the higher-order interactions once more depend on the question of necessity *versus* sufficiency in "tipping over" the threshold. All three model specifications will be estimated in the next section.

Since the dependent variable (the relative transfer decisions in the investment game) is continuous, several methods can be applied to estimate the models. A common approach in trust research is to analyze the data using OLS on raw scores (e.g. Croson & Buchan 1999, Glaeser et al. 2000, Ashraf et al. 2006, Bohnet & Baytelman 2007, among others). But in the context of experiments, which often involve a relatively low number of observations, both the presence of outliers and the (potentially non-normal) distribution of the dependent variable pose an imminent threat to the plausibility of OLS and its underlying assumptions. To account for distributional concerns and circumvent problems arising from heteroskedasticity, robust error variance estimates are usually computed when using OLS estimates. Several authors have also

used robust and weighted-least squares regression techniques to alleviate the influence of outliers (Ben-Ner & Putterman 2009, Johnson & Mislin 2011).

However, when using OLS to estimate proportions (i.e. *reltrust*, the relative amount sent), one caveat is that predictions are not guaranteed to fall inside the unit interval. In addition, the effect of explanatory variables on the predicted mean, unless they are fairly limited in their range, cannot truly be linear (Wooldridge 2002: 668). An alternative approach commonly taken in trust research to deal with the presence of "corner solution responses" is to estimate Tobit models (e.g. Fehr & List 2004, Buchan et al. 2008, Charness et al. 2008, Garbarino & Slonim 2009). The choice of a trusting act, expressed as a relative proportion, is limited to the [0,1] interval, and therefore bounded from below and above. Two-limit tobit models can appropriately deal with pileups at the endpoints of this distribution (Wooldridge 2002: 703f.), which are readily observed in the present sample as well. However, while Tobit models rest on identical assumptions about error distributions as OLS models, they are much more vulnerable to violations of those (and additional) assumptions (Maddala 1991). Data from trust experiments often involves observations at the fringes of the distribution. But in the case of an investment game context, negative values are logically implausible, because the choice of a trusting act is naturally bounded at a "zero" of distrust. Likewise, the trustor cannot more than "fully" trust, and the upper bound of one is also a natural bound, and not an effect of real censoring or truncation. At least from a logical, as opposed to a statistical standpoint, the justification of Tobit models is somewhat limited. This recommends a check with other robust estimation procedures.

Johnson and Mislin (2011) have recently proposed to use generalized linear models (GLM) to estimate decisions in the investment game context. Based on the work of Papke and Wooldridge (1996, see also Wooldridge 2002: 748ff.), they model the fractional response y, that is, the proportion or relative amount sent *(reltrust)*, as:

(1) $E(y|\mathbf{x}) = \frac{e^{\mathbf{x\beta}}}{1+e^{\mathbf{x\beta}}} = g(\mathbf{x\beta})$, or $\ln(\frac{E(.)}{1-E(.)}) = g^{-1}(\mathbf{x\beta}) = \mathbf{x\beta} + e$ (logistic link function)

(2) $Var(y|\mathbf{x}) = \sigma_{hat}^2 * g(\mathbf{x\beta}) / (1-g(\mathbf{x\beta}))$ (robust binomial variance function)

(3) $LL(\beta) \equiv \sum y_i * \ln(g(x_i\beta)) + (1-y_i) \ln(1-g(x_i\beta))$ (quasi log-likelihood function)

The link function relates the linear predictor **xβ** to the predicted values using the non-linear logistic function, which (1) ensures that the fitted values will be in the unit interval and (2) can accommodate for potential non-linearity towards the "corners" of the distribution. The results cannot be directly compared to OLS/Tobit regression, because the coefficients express the effect of a unit change on the log odds of the dependent variable. But in the present analysis, even when the coefficients are not directly comparable, so are the signs and the interaction

patterns that emerge from the estimations, and the GLM approach can provide statistical robustness-support to the other methods used.

In the following, when testing a particular model, Tobit, robust and GLM methods will be computed to verify the overall robustness of the particular specification under scrutiny. The robust procedure is based on a version of weighted least squares regression in which observations are iteratively re-weighted using calculated *Cook´s D* and residual values until the model converges to a stable estimate in which highly influential data points are downweighted, so as to alleviate their biasing influence on the parameter estimates. This is preferred to using simple OLS models, which are highly sensitive to influential data points, in particular when the number of observations is low. Non-parametric bootstrapping (2000 replications) will be used to address concerns about the non-normal distribution of *reltrust* and obtain robust variance estimates and confidence intervals. Any omitted alternative result, if not reported, can be found in Appendix A.

Another methodological note is in place at this point. The model specifications derived above demand the calculation of higher-order interactions. Concerning their statistical analysis, one concern that has been raised in the methodological literature is that interaction terms are often highly correlated to the lower-order terms by which they were formed (Cohen & Cohen 1983, Cronbach 1987, Aiken & West 1991). As a result, the interactions are closely related to the lower-order terms ("spurious multicollinearity"). The issue here is that, as multicollinearity increases (1) the predictors may explain an impressive amount of variance of the dependent variable whilst none of them is significantly different from zero, (2) the regressions may be unstable ("bouncing beta weights"), and (3) computation of the statistical models may be impossible. In essence, multicollinearity indicates that the information present in the data is insufficient to correctly allocate the variance of the dependent variable to the predictors, and it makes it difficult to distinguish the separate effects of the linear and interaction terms.

A variety of methods have been proposed in the literature to remedy this state of affairs in interaction analysis. Following the suggestions of Cohen (1978) and Aiken and West (1991), a frequently adopted solution is the "mean centering" of the lower-order terms before computing the interactions. As these authors have demonstrated, mean-centering *can* reduce the correlations between the linear- and interaction terms, and it often improves diagnostic measures of collinearity, such as the variance inflation factors (VIFs). Mean-centering also entails a second advantage in that the lower-order terms and constant can be interpreted as representing

then the conditional effects holding other variables constant at their mean. As it is, the practice of mean centering has become a standard and routine in the social sciences.[15]

However, more recent theoretical and empirical work has questioned whether mean-centering can remedy the problem of spurious multicollinearity (Kromrey & Foster-Johnson 1998, Echambadi & Hess 2007, Shieh 2011). As analytically demonstrated by Shieh (2011), mean-centering can also result in the adverse effect of *in*creasing multicollinearity among the predictors. More generally speaking, mean centering "does not change the computational precision of parameters, the sampling accuracy of main effects, simple effects, interaction effects, nor the R^2" (Echambadi & Hess 2007: 438, a conclusion that they also reach analytically). Overall, mean-centering does not substantially change the results of statistical tests of the interaction terms, and it is, if at all, advised by researchers for interpretational purposes (Jaccard & Turrisi 2003: 27f.). However, the "scaling argument" for better interpretability of the data does not apply in the present case: as it is, uncentered data provide a constant and conditional effects that pertain to the effect of a subject with "zero" accessibility; the estimated interactions pertain to the effect of increasing accessibility. Since the continuous measures are scaled to [0,1], these differences can be readily interpreted as the contrast between zero and "full" accessibility subjects. They are even more informative than the differences from the "average" subject of the sample, which a mean-centering procedure would yield.

An alternative method to deal with issues of multicollinearity that will be adopted here is that of residual centering (Lance 1988, Little et al. 2006), or "orthogonalization." The higher-order terms are first regressed on their lower-order constituents, and the empirical residuals of this regression are then used to act as the "true" and uncorrelated interaction term in the final model, containing only that part of variation which is not linearly related to the lower-order terms. As a note of caution, Echambadi et al. (2006) have shown that residual centering, while it validly assesses the true interaction effects, can lead to inconsistent estimates of the linear main effects. Thus, there is no "all-in-one" solution to address multicollinearity issues when analyzing interactions. In the following, we will make use of a combination of methods to assess the overall robustness of the particular models, relying both on residual-centered and uncentered regression estimates to tackle any potential issue of multicollinearity.[16]

---

[15] Another procedure is to use *standardized* variables to construct interactions. However, the standardization of main effect variables has another considerable impact on estimates because it involves a stochastic scaling adjustment which itself is subject to sampling error (the empirical standard deviation estimate). This can lead to wrong standard error and biased coefficient estimates. There are also additional issues of interpretation, which need not be discussed here. In general, researchers recommend against the use of this procedure (e.g. Aiken & West 1991: 42f.; Jaccard & Turrisi 2003: 68).

[16] In line with Little et al. (2006), a general empirical result of residual centering is a significant reduction of multicollinearity measures, such as the coefficient VIFs. While they were passing traditional benchmarks in a number of unorthogonal models (i.e. VIF >10 for the interaction terms), this was not the case in the orthogonal models, where none of the coefficients exceeded VIF>2.5 with exception of the three-way interaction term (the VIFs for the third-order interaction were passing beyond the benchmark of VIF>10 even after orthogonalization in most models).

6.5.2. Chronic Frame and Script Accessibility

To begin the test of the model of adaptive rationality, the three model specifications will be analyzed using multiple regression methods in the following. Focusing first on the influence of chronic script accessibility, model specification (1) is estimated and presented in table 9. In order to provide a direct assessment of robustness to the reader, the results of all three estimation methods are presented. Each model was computed separately with and without control variables. Since *recscale* and *trustscale* are scaled to the unit interval, the higher-order interactions can assume values larger than one – this is a pure scaling effect and does not carry any substantial meaning.

Table 9: Trust and chronic script accessibility

Variable	Tobit		Robust		GLM[1)]	
end	-0.891**	-0.936**	-0.736**	-0.771**	-3.197**	-3.366**
	(-2.31)	(-2.39)	(-2.34)	(-2.33)	(-2.31)	(-2.31)
frame	0.084	0.007	0.094	0.046	0.401	0.129
	(0.19)	(0.02)	(0.23)	(0.11)	(0.27)	(0.08)
recscale	0.0741	0.151	0.084	0.155	0.328	0.554
	(0.2)	(0.36)	(0.25)	(0.41)	(0.25)	(0.37)
*end*rec*	1.184**	1.253**	0.937**	1.001**	4.168**	4.427**
	(2.05)	(2.15)	(1.99)	(2.03)	(2.05)	(2.07)
*frame*rec*	-0.077	0.042	-0.109	-0.035	-0.456	-0.039
	(-0.12)	(0.06)	(-0.18)	(-0.06)	(-0.21)	(-0.02)
*end*frame*	1.143*	1.179*	0.926*	0.857+	3.777*	3.841*
	(1.73)	(1.84)	(1.65)	(1.53)	(1.66)	(1.65)
*end*frame*rec*	-1.778*	-1.820*	-1.437*	-1.336+	-5.864*	-5.960*
	(-1.83)	(-1.93)	(-1.76)	(-1.63)	(-1.77)	(-1.76)
trustscale	0.399**	0.310+	0.352**	0.281*	1.345**	1.118+
	(2.08)	(1.46)	(2.31)	(1.67)	(2.16)	(1.63)
constant	0.203	0.187	0.219	0.207	-1.072	-1.218
	(-0.71)	(-0.54)	(-0.89)	(-0.7)	(-1.10)	(-0.99)
Pseudo R^2 (ps. LL)	0.051	0.084	0.07	0.11	(-157.8)	(-155.18)
Wald (full model)	20.66***	36.28***	27.22***	44.05***	20.04***	33.76***
χ^2 Improvement (4df)	8.4*	7.9*	9.12*	7.38+	8.17*	7.23+
Control variables	No	Yes	No	Yes	No	Yes

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. [1)] Effects on log-odds. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

The three methods provide a consistent picture of the interaction pattern emerging from the first model specification. Moreover, when computing the residual-centered, orthogonalized models to assess the robustness of the interactions and ensure against spurious multicollinearity, the results are identical to those obtained above (see Appendix A). Overall, this provides a great deal of confidence that the results tap on a substantial relation among predictors and independent variables and not on statistical artifacts of some sort.

The estimated interaction pattern is matching with predicted interaction pattern number two (see section 6.3.3), which admits a negative three-way interaction and a positive two-way interaction of *end*frame*, as well as *end*rec*. In other words, when varying chronic script accessibility and the experimental factors, those variables working in opposition to cognitive motivation (high accessibility *or* a cooperative framing condition) have each been *sufficient* to reduce the negative effect of high initial endowments on unconditional trust. The negative three-way interaction is reaching marginal conventional statistical significance in all models (for example, t=-1.93, p=0.054 in model 2). The joint contribution of the four interaction terms is acceptable and improving model fit, as compared to a situation in which they are assumed to be zero (Wald tests between $\chi 2$ (4) =9.12, p=0.058 in model 3 and $\chi 2$ (4) =7.23, p=0.1224 in model 6).[17] A direct comparison of the estimated coefficients between the Tobit and robust regression methods reveals that the Tobit slopes are steeper, potentially reflecting the difference in how the models deal with the corner solutions present in the data. None of the control variables has a noticeable effect on the level of *reltrust* (see Appendix A), and their introduction does not substantially change the results.

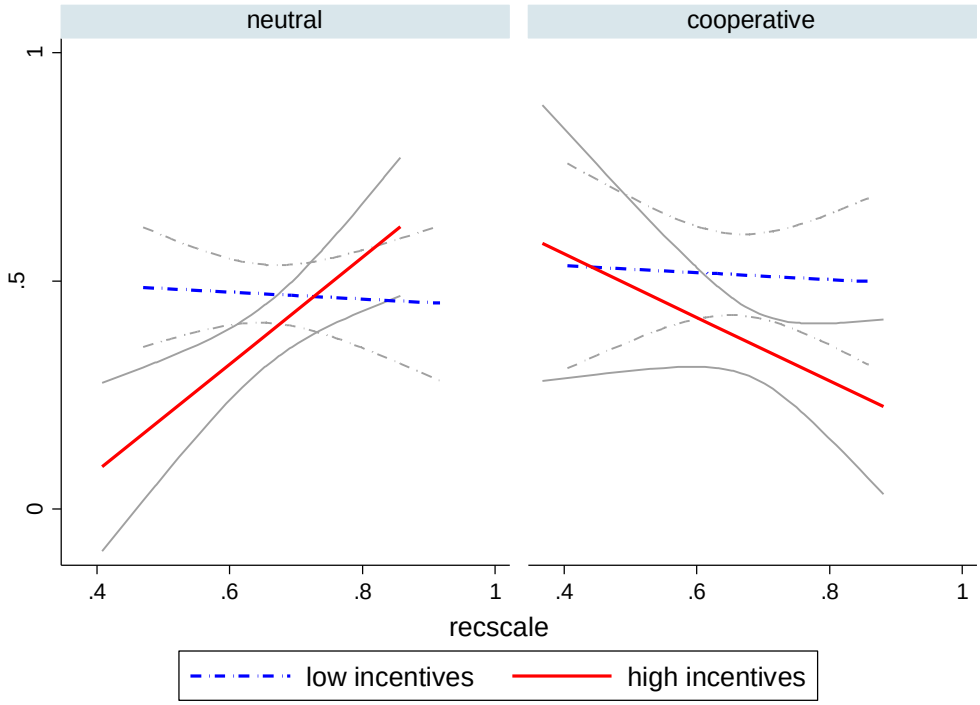
Most importantly, the models uncover a negative incentive effect which is counterbalanced by high script accessibility. This effect is present in the neutral framing condition. In other words, trustors in fact trust less and switch to conditional trusting strategies when the "stakes are high," but this effect can be overrun by high chronic accessibility of a trust-related script. Only those subjects scoring low on the norm of reciprocity scale do in fact respond to the incentive treatment as expected under main hypothesis H4a. With increasing norm internalization, the negative effect of instrumental incentives on cognitive motivation diminishes. This lends support to hypothesis H8 (and qualifying H4a), stating that the effect of cognitive motivation is mediated by chronic script accessibility. The result adds an interesting twist to the experimental-economic investigation of "stake size" effects (Camerer & Hogarth 1999): as suggested by the model of adaptive rationality, they cannot be appropriately accounted for without regard to relevant frames and scripts, because a high internalization (of norms, roles, rules, routines etc.) may fully suppress such incentive effects. This is what can be observed in the neutral framing condition. The model also uncovers a strong and significant effect of trust-related frames (*trustscale*) when estimating model specification (1).

At the same time, incentive effects are no longer present in the cooperative framing condition, as indicated by the negative three-way interaction and the positive two-way interaction of

[17] Testing the null-hypothesis of a joint zero effect of the higher-order interactions when the frame accessibility measure (*trustscale*) and controls are excluded improves the test´s results and does not change the predicted interaction pattern. In this case, both the Tobit and robust regressions estimates deliver a Wald-test on the joint effect of the interaction terms which predicts a non-zero effect with p<0.05 (results omitted).

*end*frame*. Principally, the models lend support to hypothesis H9, according to which the cooperative framing condition increases the activation weights of trust-related frames and scripts and mediates the impact of incentives and motivation on the degree of rationality during interpretation and choice. That is, a switch to conditional trusting strategies in the face of high stakes can also be prevented by a cooperative framing of the trust problem, and by a presence of situational cues which indicate the validity and appropriateness of a corresponding frame or script. This finding is important insofar as it suggests that "context" and "incentive structure" do not influence the definition of the situation completely independent of each other, even when this assumption is regularly made in experimental economic and social-psychological research. What is more, the three-way interaction between chronic accessibility, situational cues and incentives supports hypothesis H10, stating higher-order interactions between all parameters. The estimated interaction pattern also conforms to one of the predicted interaction patterns. To further aid the interpretation of the statistical results, the following figure shows the predicted levels of *reltrust* for the neutral and cooperative framing conditions, separated by high and low initial endowments (figure 23):[18]

Figure 23: Predicted level of *reltrust* across experimental treatments



Focusing on the neutral framing condition, as presented in the graph on the left of figure 23, a difference in the predicted level of trust between the low- and high incentive treatment is

[18] The graphs have been constructed using the predicted values and standard errors from a Tobit model estimating model specification (1) without control variables and using a two-sided type-1-error probability of σ=0.10.

clearly visible. Adding confidence intervals around the predicted mean, it is apparent that the level of trust significantly drops with high initial endowments for subjects scoring in the bottom range of *recscale.* This incentive effect disappears with increasing chronic script accessibility, and in the upper range of *recscale*, it is not present anymore. In contrast, when focusing on the cooperative framing condition, as depicted in the right graph of figure 23, no such incentive effect is visible in the lower range of *recscale*, indicating that the cooperative framing has been equally sufficient in suppressing incentive effects for low accessibility subjects. In fact, the predicted levels of *reltrust* do not significantly differ between incentive conditions over the whole range of chronic script accessibility, as indicated by the confidence bounds.

At the same time, the models predict a negative slope for *recscale* in the cooperative framing condition when the "stakes are high." This finding contradicts hypothesis H7, according to which the effect of situational cues on the activation weight varies positively with chronic frame- and script-accessibility (and *vice versa*). As stated in H7, context effects should be more pronounced for high-accessibility subjects. In the present data, high accessibility subjects do *not* react to the framing treatment with relatively *more* trust, whereas low accessibility subjects do. Comparing between framing conditions and holding incentives constant, a small level effect of neutral *versus* cooperative framing is visible in the low endowment conditions; but the positive relation between *reltrust* and *recscale* is reversed with high initial endowments. It is important to keep in mind that the models test the overall interaction pattern emerging from the data, and not exclusive separate main- or interaction effects. As it is, interaction pattern number 2 predicts *recscale*frame* to have a coefficient that is smaller or equal to zero, which is what we observe in the regression results.

As reported in the descriptive statistics, the average level of *reltrust* in the high/cooperative condition (M=0.37) is the lowest of all four factorial constellations, and it is particularly low when compared to the conditional mean in the low/cooperative condition (M=0.50). Even when the differences in trust appear to be insignificant from a comparison of means and their confidence intervals, as based on the multiple regression models, a simple test of conditional means *within* the cooperative framing condition finds that *reltrust* is lower in the high incentive condition (two-sided Wilcoxon rank-sum test, Z=2.297, p=0.022).[19] Thus, the predicted negative slope is probably more than a statistical artifact from the estimations and points towards a more substantial finding that warrants explanation.

[19] This result could also be addressed in terms of random sampling error. Concerning this possibility, note that the observed mean differences between high- and low endowments in the cooperative framing are large and result in an estimated total effect size of d=0.41. Given that the Wilcoxon rank-sum test referred to above was conducted using a two-sided type-1-error confidence level of α=0.05, it is very unlikely that the rejection of the null-hypothesis was made erroneously, even when we cannot exclude this alternative with certainty.

The finding that context and incentive structure interact is not surprising. It is a direct implication of the model of adaptive rationality. The results conform to the predictions made in generic dual-processing approaches, which suggest a *de*creasing influence of chronically accessible knowledge with *in*creasing situational strength, and *vice versa* (Mayerl 2010: 42). More generally speaking, the DP models predict accessibility effects of primed constructs (Higgins 1996). This is what we observe in low-accessibility subjects: in the face of high initial endowments *and* cooperative cues, but in the lack of an internal regulative script, they respond to the trust problem by using the contextual information to adjust their trusting strategy. This might be termed a "mere priming-effect" (the analysis of decision times further helps us to understand the observed behavior in terms of controlled *versus* automatic processing, see below). On the other hand, high accessibility subjects are sensible to whether or not a cooperative framing is presented in conjunction with high- or low initial endowments, and from the results, one cannot exclude the possibility that they switch to conditional trusting strategies in the high/cooperative condition. In fact, a simple test of central tendency reveals a significant drop in *reltrust* for the high accessibility subgroup between endowments in the cooperative framing condition (WRT, Z=1.978, p=0.0479), while no such effect can be found for low accessibility subjects (WRT, Z=1.245, p=0.2133).

One plausible explanation for this finding is that high accessibility subjects, in contrast to low accessibility subjects, experience a "mismatch" in that a cooperative context *and* high initial endowments collide. This does not imply that high endowments produce a separate symbolic cue independent of the framing condition for all subjects. If this was the case, then the high/neutral condition would not reveal the neutralizing effect of chronic script accessibility, and a consistent negative incentive effect across all framing conditions and for all accessibility groups would be observed. In contrast, the data indicate that a "mismatch" is contingent upon being high in chronic accessibility, so that an attribution of symbolic meaning to the "cue" of high endowments, if at all, has been made by trustors in the high accessibility group only. But are there any theoretical arguments that support such an assertion?

To begin with, it is unlikely that the chronic accessibility of frames and scripts does *not* influence other parameters of the mode-selection threshold. But this is implied in the current MFS formalization by treating the effect of "external" sources of variation, such as the presence of situational cues o_i, as independent from accessibility, and *vice versa*.[20] A consistent finding in social-psychological research is that chronically accessible knowledge structures can also in-

[20] According to Kroneberg (2011: 130), the parameter of situational cures o_i is an *objective* measure of the presence (or non-presence) of *significant* situational objects, which indicate the appropriateness of a particular frame in the current situation. One important assumption here is that "significance" can be objectively ascribed (given that frames are socially shared and "objectified," including shared definitions of significant cues which indicate their appropriateness). The following argument opens up the interesting question of whether "significance" can ever be objectively identified independently of perceiver characteristics.

voke implicit goal-setting (Bargh et al. 1996, Aarts & Dijksterhuis 2000). One of the earliest demonstrations of such automatic motivational effects is the finding of a heightened "perceiver readiness," that is, of a selective perception of, and attention to, cues which are related to the chronically accessible constructs (Higgins et al. 1982, Bargh & Pratto 1986). While the model of frame selection models a one-way causal path from cues to activation weights, treating o_i as an "external" source of variation in the threshold, one can argue that $o_i = o_i(a_i, a_j)$, that is, the *subjective* presence of significant situational cues which actors perceive is also a function of chronic accessibility. In effect, this implies selective perception, attention, and perceiver readiness. Note that this cannot be covered in the link l_i between knowledge and objects either, as this parameter describes an invariant property of stored knowledge structures, related to the strength of the symbolic relation between objects and cues. Thus, it could be possible that low-accessibility subjects are *not* selectively attentive to the cooperative cues of the context (even when this subgroup can use the cues in the sense of "temporary accessibility" and priming), while this is the case for high-accessibility subjects.

Secondly, if the activation of trust-related frames and scripts does not only rely on the presence and appropriateness of relevant significant situational cues, but also on the *absence* of distractions, a "mismatch" can arise if the generalized trust-related frames and scripts do *not* unconditionally extend to high-cost situations *per se*, and if high initial endowments represent a nuisance to a "favorable" subjective definition of the situation, once adopted. Note that the visual cues and neutral/cooperative wording manipulations were presented *before* presenting the investment game instructions, and therefore *before* presenting the incentive manipulation. This specific design-feature was implemented because the general instructions referred to "other" participants already. To keep a consistent terminology throughout the experiment, the wording manipulation (Partner/Team *versus* Participant/Group) was already used from the very beginning of the experiment. Assuming that the framing manipulation was successful (as indicated by the behavior of low accessibility subjects), and assuming that the activation of a trust-related frame by high-accessibility trustors directs selective attention towards confirming and disconfirming cues, a "high cost cue" could have presented a threat to their pre-established subjective definition of the situation. High initial endowments then would acquire a symbolic meaning of nuisance for high accessibility subjects, indicating the *in*validity of the currently adopted frame to them, but not affecting the low accessibility group.

In fact, alternative specifications of the match (Esser 2001: 269ff., Stocké 2002: 127ff.) have been proposed in which the *absence* of situational nuisances, distractions and interruptions has been included as a separate and independent factor determining the match, such that $m_i = a_i * o_i * e$, where e_i [0,1] describes the absence of nuisance. If high endowments have presented a nuisance (the analysis of decisions times in the next chapter supports this assertion), then we need to conclude that the framing manipulation had an unexpected effect for high-

accessibility subjects, but we can also extract information and draw important conclusions from this observation, in that the current experimental data *support alternative formulations of the match* which can accommodate for separate effects of (1) significant situational objects and (2) situational nuisances. This issue will be further scrutinized when exploring subgroups of the sample and analyzing decision times.

Focusing now on the effect of chronic frame accessibility and its interactive effects on *reltrust*, the regressions using model specification (2) establish a different result. Principally, while a simple main effect of *trustscale* can be found when analyzing a model without interactions, no interaction pattern can be estimated with sufficient statistical certainty in the full interactive models. None of the coefficients is interpretable, and neither the introduction of control variables nor the computation of orthogonal models substantially changes this result (table 10, omitted results reported in Appendix A):

Table 10: Trust and chronic frame accessibility

Variable	(1) Tobit	(2) Tobit	(3) Tobit	(4) Robust	(5) GLM[1)]
end	-0.124***	0.143	0.089	-0.095	-0.114
	(-2.71)	(0.43)	(0.26)	(-0.34)	(-0.10)
frame	-0.001	0.171	0.167	0.026	0.404
	(-0.03)	(0.43)	(0.42)	(0.07)	(0.32)
trustscale	0.418**	0.491	0.433	0.232	1.121
	(2.17)	(1.43)	(1.2)	(0.76)	(0.95)
*end*frame*		-0.517	-0.339	-0.123	-1.067
		(-1.02)	(-0.66)	(-0.28)	(-0.64)
*end*trustscale*		-0.393	-0.301	0.001	-0.392
		(-0.71)	(-0.54)	(0)	(-0.21)
*frame*trustscale*		-0.249	-0.231	0.001	-0.517
		(-0.37)	(-0.34)	(0)	(-0.24)
*end*frame*trust.*		0.77	0.483	0.122	1.444
		(0.9)	(0.56)	(0.17)	(0.52)
recscale	0.167	0.167	0.296	0.292	1.066
	(0.69)	(0.68)	(1.01)	(1.15)	(1.04)
constant	0.145	0.089	0.004	0.0981	-1.675+
	(0.7)	(0.33)	(0.01)	(0.35)	(-1.50)
Pseudo R^2 (ps. *LL*)	0.028	0.032	0.065	0.095	(-156.54)
Wald (full model)	10.93**	13.75*	22.31*	29.15***	22.4**
χ^2 Improvement (4 df)	-	1.54	0.75	0.55	0.81
Control Variables?	No	No	Yes	Yes	Yes

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. [1)] Effects on log-odds. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

The first column assesses main effects only. It predicts a negative main effect of the endowment treatment (t=-2.71, p=0.007) and a positive main effect of chronic frame accessibility (t=-2.17, p=0.03) on the observed level of trust, holding *recscale* constant. No effect of the framing treatment can be found. With respect to main effect hypotheses, this lends support to

hypotheses H1 and H4a, but none to H3. That is, (1) unconditional trusting strategies are more likely with higher chronic accessibility of a trust-related frame, (2) increasing cognitive motivation *via* initial endowments pushes behavior towards the distrust equilibrium, indicating a prevalence of conditional trusting strategies, but (3) the framing of the trust problem does not exhibit any detectable main effect. The coefficient of *trustscale* is similar to the one obtained in the first model specification. Since *trustscale* is limited to the unit interval, its coefficient expresses the difference between a hypothetical zero and a "full" chronic frame accessibility subject. The predicted difference in *reltrust* between zero- and full frame-accessibility is considerable with about a 40 per cent difference.

The model in the second column assesses the interaction pattern emerging from the interplay between initial endowments, framing condition and chronic frame accessibility. While the overall explanatory power of the model seems to be weakly better than that of a null-model ($\chi 2$ (7) =13.75, p=0.085, Tobit regression), the inclusion of the interaction terms adds no explanatory power to the model ($\chi 2$ (4) =1.54, p=0.81), and the test cannot reject the hypothesis that the interaction terms are jointly zero. Overall, the model and coefficients are not interpretable. This result does not change when including control variables (column 3). Neither *append, age, sex, partner*, *fiscale* or *nfcscale* have an effect that is estimated anywhere near statistical certainty. We observe similar results when using a robust regression method (column 4) or when estimating the model specification with the GLM approach (column 5). The only marked difference arising from computing the residual-centered and orthogonalized models is that a negative main effect of the initial endowment condition is now consistently estimated, revealing a significant 11 per cent decrease in *reltrust* across framing conditions. This difference corresponds to the empirical raw mean difference between endowment conditions. All other coefficients remain insignificant (see Appendix A).

There are several potential explanations to help us understand this result. On the one hand, the effects captured in *trustscale* and emanating from the chronic accessibility of a trust-related frame may simply be very weak because what we actually observe in the experiment is a choice, and not interpretation. Thus, any effect of chronic frame accessibility might have already played out its part before actions are observed. Then, the variable *trustscale* would be less important to modeling the observed choice of a trusting act. As suggested by the model of trust and adaptive rationality, the mode-selection threshold in the stage of action selection is tied to more stringent conditions. An important determinant of action selection is chronic script accessibility, which comes into play only after the stage of interpretation has been completed. On top of that, the norm of reciprocity is a highly regulative script. Thus, model specification (2) might tap on effects which are too weak and too small in effect size to be reliably detected by the current data set and a limited number of observations. In fact, *trustscale* was

found to exert a significant influence when estimating model specification (1), where the focus was on the interplay between script accessibility and experimental treatments.

On the other hand, *trustscale* might interact with *recscale* in determining the match, as suggested by model specification (3). If this is the case, and if script accessibility is equally important in determining action, then the above model could simply be insufficiently specified and important variables are omitted. Furthermore, it may be that the observations are heterogeneous in their response to the framing treatment, depending on other unobserved characteristics. Thus, similar to a neglect of *recscale*, other variables might influence the interaction pattern in a way that prevents the model from capturing the true effect of *trustscale*. Lastly, the measure of generalized trust, as captured in the short version of the "Interpersonal Trust Inventory" (Kassebaum 2004), might simply not provide a relevant frame for the experimental setting. While a partial answer to this issue can be given when analyzing model specification (3), an alternative specification that explores the possibility of subgroup heterogeneity will be explored in section 6.7. Suffice it to say at this point that the third possibility, a complete irrelevance of trust frames, is not supported by the data.

Model specification (3) addresses the joint effect of chronic frame and script accessibility *and* situational cues on the activation weight, holding the effects of cognitive motivation constant. Since we now introduce interactions between two continuous variables that are scaled to the unit interval, the magnitude of coefficients can become much larger, but this again is a mere scaling effect. Estimating model specification (3) reveals following results (table 11):

Table 11: Trust and the activation weight components

Variable	Uncentered			Orthogonal		
	Tobit	Robust	GLM[1)]	Tobit	Robust	GLM[1)]
frame	-2.398	-1.471	-6.44	-0.004	-0.01	-0.022
	(-1.26)	(-0.95)	(-0.97)	(-0.09)	(-0.27)	(-0.14)
trustscale	-4.110**	-2.723*	-11.46+	0.421**	0.382***	1.441**
	(-2.01)	(-1.83)	(-1.64)	(-2.15)	(-2.59)	(-2.27)
recscale	-2.996*	-1.917+	-8.059	0.167	0.126	0.608
	(-1.71)	(-1.46)	(-1.34)	(-0.67)	(-0.6)	(-0.71)
*frame*trustscale*	5.206+	3.364	14.55	-0.028	0.023	-0.01
	(-1.61)	(-1.31)	(-1.32)	(-0.07)	-0.07	(-0.01)
*frame*recscale*	3.547	2.13	9.427	-0.991**	-0.831**	-3.316**
	(-1.3)	(-0.96)	(-0.99)	(-2.06)	(-2.00)	(-1.97)
*trustscale*recscale*	6.545**	4.432**	18.54*	2.987	1.63	5.329
	(-2.19)	(-2)	(-1.8)	(-0.98)	(-0.69)	(-0.54)
*frame*trust.*rec.*	-7.726*	-4.933	-21.47	-1.493	0.74	0.512
	(-1.65)	(-1.34)	(-1.35)	(-0.30)	(-0.18)	(-0.03)
end	-0.119***	-0.120***	-0.470***	-0.120**	-0.126***	-0.461***
	(-2.60)	(-2.98)	(-2.92)	(-2.56)	(-3.10)	(-2.93)
constant	2.349*	1.628*	4.727	0.142	0.184	-1.272*
	(-1.96)	(-1.81)	(-1.15)	(-0.69)	(-1.06)	(-1.83)
Pseudo R^2 (ps. *LL*)	0.0858	0.111	(-155.39)	0.0429	0.0636	(-158.53)
Wald (full model)	33.31***	38.13***	29.11***	17.63**	23.15***	17.49**

χ^2 Improvement (4 df)	7.87*	7.04+	6.2	5.2	5.11	4.19
Control Variables?	Yes	Yes	Yes	No	No	No

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. [1)] Effects on log-odds. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

When estimating the third model specification, differences in the regression results can be observed between the three estimation methods. While the Tobit model suggests a considerable interplay between all of the activation-weight determinants, this result is estimated with less statistical certainty in the robust and GLM approach, even when the predicted signs of the coefficients do not differ and the *trustscale*recscale* interaction remains significant. The magnitude of effects is considerably lower in the robust approach. The Wald tests for a joint non-zero effect of the interaction terms become less supportive, and they are least optimistic in the GLM. What is more, in the case of model specification (3), the computation of the orthogonalized models, as presented on the right of table 11, suggests that the correlations between the predictors and IV/DV´s may be spurious, and that the uncentered models may suffer from biased coefficients and inflated standard errors. Thus, we cannot exclude the possibility that the uncentered models are biased because the higher-order interactions between the two continuous measures have introduced spurious multicollinearity. Overall, this warrants caution when judging the models. Interpretation of the coefficients in the uncentered model specification (3) will not be further pursued here.

At the same time, the orthogonal models consistently uncover a significant influence of two predictor variables. First, the conditional main effect of *trustscale* is now estimated similarly to model specifications (1) and (2), predicting an average difference of about 40 per cent in *reltrust* across endowment and framing conditions for a hypothetical zero *versus* full frame accessibility actor. Thus, even when model specification (2) did not uncover an interactive pattern between treatments and trust-related frames, a conditional main effect can be established. Second, a negative interaction between *frame* and *recscale* is uncovered. This finding supports the "mismatch" hypothesis as stated above. Obviously, high chronic script accessibility subjects trust relatively less in the cooperative framing condition than their low accessibility counterparts. Including the omitted control variables does not change the results substantially (see Appendix A). Overall, while the standard approach would suggest a considerable amount of interplay between the two activation-weight determinants, this result is not stable and should be taken with a grain of salt. The orthogonal approach, on the other hand, does not add any new information. It more directly reveals the *frame*recscale* interaction, which the previous model specifications have already suggested, and confirms a conditional main effect of *trustscale*.

6.5.3. NFC/FI as Mode-Selection Determinants

Up to this point, the variables *nfcscale* and *fiscale* have been held constant and used as control variables when analyzing the influence of chronic accessibility measures and their interplay with manipulations of the context and the incentive structure of a trust problem. However, both constructs can be justified to exert an influence on the mode-selection threshold in their own right. We can think of them as further determinants of the mode-selection threshold, even when they have not been incorporated directly into the theoretical conceptualization of the model of adaptive rationality so far.

First, the need for cognition (NFC) of the individual actor can be regarded as an individual-intrinsic, as opposed to situational-extrinsic, aspect of cognitive motivation. A parameter of intrinsic cognitive motivation can be easily included in the mode-selection threshold by extending the model. Assume that the selection of the rational mode does not only incur certain processing costs C, but also affords some intrinsic utility U_{int} which reflects the actor´s preference for adopting a rational processing mode and the corresponding "joy of thinking." Thus, in deriving the mode-selection threshold, an additive component U_{nfc} can be introduced that counter-balances the inhibitive effect of cognitive processing costs C. Modeling the states of the world of the rational mode, the certain consequences of its activation then include $(C+U_{nfc})$ instead of merely C.

Second, faith in intuition (FI) can be incorporated with a similar extension of the model. As noted by Pacini & Epstein, the FI scale was designed to represent the "intuitive-processing counterpart" (1999: 973) of the NFC scale. It contains information about the intrinsic utility from relying on intuition and captures the preferences that an actor has towards automatic processing. Assume that the selection of the automatic mode includes the additive utility component U_{fi}, indicating the constant intrinsic utility stemming from a preference for intuition. Straightforward (but tedious) algebra yields that the U_{fi} measure will be located at the same position as its NFC counterpart with an opposite sign. Thus, the extended version of the mode-selection threshold, including processing preferences, and solved here for the automatic selection of a frame, can be formulated as:

$$m_i > 1 - (C + U_{fi} - U_{nfc}) / (p * (U_{rc}+C_w))$$

This implies that U_{fi}, the preference for intuitive processing, affects the mode-selection threshold in favor of the automatic mode (it works in the same direction as C), while its counterpart U_{nfc}, the need for cognition, does the opposite (it reduces the inhibitory influences of C and U_{fi}). All three terms are derived as additive, implying that either of them can be *sufficient* to exert a "tip-over" influence on the mode-selection threshold. Moreover, this implies that all three parameters exert an influence that is independent of each other (for NFC/FI, this proposition has received empirical support, see Epstein et al. 1996, and Pacini & Epstein 1999).

As usual, the effect of the threshold parameters is expected to change in interaction with the other parameters of the mode-selection threshold (H10). It is straightforward to derive interaction patterns for the variable U_{fi} when varying the experimental factors simultaneously. To see why this is the case, note that the effect of a change in either C or U_{fi} on the balance of the mode-selection threshold is similar to the effect of a change of one of the left-hand side parameters such as chronic accessibility or situational cues, that is, increasing accessibility *or* increasing processing costs both pushes the threshold in favor of the automatic mode. In particular, if we are interested in the interplay of U_{fi} with the experimental factors, we can directly replace the optimal thresholds a* (which trigger a "tip-over" for chronic accessibility measures) with optimal U_{fi}*, and analyze a corresponding model in which either $U_{fi}^* < U_{fi}$ ($=U_{high}$) or $U_{fi}^* > U_{fi}$ ($=U_{low}$). When comparing the values of the right-hand side to the left-hand side of the threshold and varying the framing treatment o_i on a low or high level (the principal setup is similar to the approach presented in Appendix C), the derived interaction patterns are equal to those presented in section 6.3.3. In the case of NFC, the sign of effects is opposite to that of U_{fi}. Thus, the predicted interaction patterns are still the same, but their signs are exactly opposite to those predicted before. Empirically, two model specifications will be analyzed:

(4) $E(reltrust|\mathbf{x}) = \mathbf{x\beta} + e = \beta_0 + \beta_1 * end + \beta_2 * frame + \beta_3 * nfcscale + \beta_4 * end * frame + \beta_5 * frame * nfcscale + \beta_6 * end * nfcscale + \beta_7 * end * frame * nfcscale + fiscale$ + controls +e

(5) $E(reltrust|\mathbf{x}) = \mathbf{x\beta} + e = \beta_0 + \beta_1 * end + \beta_2 * frame + \beta_3 * fiscale + \beta_4 * end * frame + \beta_5 * frame * fiscale + \beta_6 * end * fiscale + \beta_7 * end * frame * fiscale + nfcscale$ + controls +e

The control variables then include *trustscale* and *recscale*, to hold the other mode-selection threshold parameters constant. When empirically testing both model specifications, the results are unambiguous and allow for one simple conclusion: none of the variables *fiscale* or *nfcscale* does interact with the decision to trust in model-specifications that include interactions between the processing preferences and the experimental factors (results omitted, see Appendix A). In the case of the framing treatment, the above analysis does not reveal any differences between high or low NFC/FI individuals in the susceptibility to framing effects, and with respect to the incentive manipulation, there is no indication that the NFC/FI processing preferences independently influence the response of the trustors to high or low stakes.

Computing residual-centered orthogonal models does not reveal any new information or improve the estimates either, even when an endowment effect of *end* of about minus 11 per cent across framing conditions is revealed (this matches with the results from the analysis of accessibility measures in section 6.5.2). Apart from that, none of the model coefficients can be con-

fidently interpreted, and we cannot reject the hypothesis that processing preferences, as measured by the REI scales, do not directly affect the choice of a trusting act.

It is important to note, however, that processing preferences do not provide any "substantial" trust-related information to a trustor regarding when defining a trust problem and deciding about the choice of a trusting act. They are, so to say, a "context free" preference for automatic versus rational processing. Thus, in the context of a trust problem, the REI scales may be more influential in how available information is dealt with and how accessible information is used, but they do not determine *which* information actors will attend to, and "what comes to mind". This issue will be taken up again in section 6.7.3, where models that vary processing preferences *and* chronic accessibility simultaneously will be analyzed. As presented below, these models suggest that an analysis of trust must take care of accessibility and processing preferences, a finding which is also reasonable from a theoretical standpoint.

6.5.4. Discussion

Overall, a number of important conclusions can be drawn from the analysis of model specifications (1)–(5). Firstly, the models estimate an overall negative endowment effect across framing conditions with about an 11 per cent drop in *reltrust* in specifications (3) to (5). This corresponds to the empirical mean difference which was found in the descriptive statistics, and confirms a general main effect prediction: when much is "at stake," cognitive motivation increases and pushes the trustors towards a rational consideration of the trust problem and into conditional trust. However, this finding must be qualified. As suggested by model specification (1), trustors exhibit a differential response to an increase in cognitive motivation depending both on the chronic accessibility of a trust-related script and on the context. The interaction pattern revealed in model specification (1) is consistent with one of the predicted MFS interaction patterns. High norm internalization can suppress the effect of instrumental incentives which push actors toward a rational consideration of the trust problem. Thus, even in the face of high stakes, trustors can choose an unconditional trusting strategy if relevant scripts are chronically accessible, *or* if the context supports a favorable definition of the situation. Moreover, when estimating this model, a conditional main effect of trust-related frames is also revealed. In an unexpected twist, high accessibility subjects (in contrast to low accessibility subjects) were found to react with a more conditional trusting strategy and lower levels of trust when a cooperative context *and* high initial endowments were combined. It is important to discern whether this finding only challenges the assumptions which were made in designing the current experiment, or whether it bears a more substantial meaning that needs to be addressed theoretically in terms of the model of adaptive rationality and modeling of the mode-selection threshold.

On the one hand, it may be the case that the assumptions made in designing the experiment were inadequate. In particular, this questions the overall efficiency of the framing treatment and casts doubt on whether the incentive treatment had an exclusive effect on cognitive motivation. To begin with, the framing manipulation was designed to present relevant situational cues to the subjects and influence the match m_i of trust-related frames and scripts. A naïve conclusion from analyzing the simple (unconditional) main effect of the treatment variable *frame* is that it simply had no effect. However, it was also found that low (and high!) accessibility subjects readily use the presented cues to adjust their trusting strategies. Thus, it would be erroneous to conclude that there is no effect of the presented context, even when the effects are conditional. Secondly, it can hardly be argued that the incentive manipulation has emitted a negative symbolic cue across all conditions and for all subjects. If this were the case, then we should not have observed the balancing function of script accessibility in the neutral framing condition, and we should not have observed an effective priming of low accessibility subjects in the high/cooperative condition. Both observations contradict the general "stakes-as-a-symbol" hypothesis. Put sharply, it is only the *high* accessibility group whose behavior departs from the model´s predictions in the high / cooperative condition. A symbolic effect of high initial endowments could have been present for this subgroup in this experimental condition, but why?

The explanation favored here is that of a "selective mismatch" and situational nuisance which emerges exclusively in high accessibility subjects. The argument invokes selective attention and holds that high initial endowments may serve as a salient cue for (selectively attentive!) high-accessibility subjects, who do not further only rely on their (successful!) initial definition of the situation with a trust-related frame and script, but who are also attentive to whether their interpretation and automatic application of stored knowledge is still correct. High stakes only "fit the frame" if their presence is encoded as a typical situational element of generalized trust frames and scripts. Arguably, this bridge hypothesis *cannot be tested with the current data*, and we need to rely on indirect evidence.[21] First, model specification (3) revealed a negative *frame*recscale* interaction which exactly pin-points this adverse effect. Second, model specification (1) showed that the cooperative cues were readily used by low accessibility subjects in the face of high endowments, suggesting a "priming effect" for this group. Overall, the data lend some plausibility to the "mismatch" explanation. Further evidence will be presented when analyzing the decision time data, which show a decision time increase for high accessibility subjects in this experimental condition, suggesting a switch to more elaborate and conditional trusting strategies.

[21] Of course, one could argue that, from a normative standpoint, this is precisely what the norm *typically* should prescribe: unconditional trust irrespective of the "stakes". But, empirically, the question remains whether the real-life actors (the subjects participating in the experiment) have learned and acquired a norm that includes high-cost situations as a "typical area of application" for trust-related norms and scripts.

As a theoretical consequence, this encourages a consideration of conceptualizations of a match m_i in which nuisances are taken care of. A similar argument was also made by Mayerl (2009: 235), who adopts am earlier conceptualization put forward by Esser (2001: 270) and models the match as $m_i = ai*o_i*e_i*$, where e_i [0,1] represents the absence of nuisance. In fact, Kroneberg (2005: 351) also states that the activation weight can be reduced by a factor (1-d), where d [0,1] represents the presence of nuisances. According to Kroneberg, this factor should be introduced into the analysis "on demand." In the current work, this seems to be the case. Taking a more general stance, the present data suggest that the (non-)emergence of a subjective nuisance is strongly dependent on characteristics of the actors, for example, the accessibility of stored frames and scripts. This kind of *reflexive feedback between stored knowledge structures and "on-line" cognition* has not been theoretically incorporated and dealt with yet. Principally, by modeling o_i as $o_i(a_i, a_j)$, one can invoke selective attention and perceiver readiness. This idea merges well with the dynamic and interactive conceptualization of a social construction of trust, as put forward in the previous chapters; it directs our attention to the possibility that the modeling of adaptive rationality can be complicated by endogenous processes such as a reflexive feedback between "active" cognition and stored knowledge structures.

Model specifications (2), (4) and (5) have each tested the effect of one continuous measure (frame accessibility or either of the two processing preferences) on *reltrust*, including interactions with the two experimental factors. While neither of the models reveals substantial interactive effects of the variable under scrutiny on the level of trust, it is argued here that the conclusion of a "failure" of the adaptive rationality model would be premature. If substantial relations among the threshold variables are not included, the models can be miss-specified and result in weak detected effects and poor estimation results. Concerning the effect of processing preferences, it is likely that they do not influence the choice of a trusting act independent of "what comes to mind". Therefore, an analysis of subgroups and an estimation of models in which processing preferences and accessibility measures vary simultaneously will be conducted further below. The same holds true for the finding that no main effect influence of *trustscale* was found. Either, the effect of script accessibility and its interactions with the other mode-selection determinants is stronger (this is plausible given that we observe a choice and not interpretation) or its effect is masked by heterogeneity which can only be uncovered when analyzing subgroups. In fact, a conditional main effect of *trustscale* could be revealed in model specifications (1) and (3).

Arguably, one severe limitation of the current data set is the low number of observations. Given that models estimate higher-order interactions, a larger sample size would have been desirable. For example, Kroneberg (2011c) uses Monte-Carlo simulations to assess the statistical

testability of the MFS interaction hypotheses and concludes that approximately N=2000 observations constitute the optimal sample size.

A methodological corollary that can be drawn from the present analysis is that a combination of different statistical methods can support interpretation and provide a robustness check of results in trust research. Importantly, the Tobit models have, in several cases, detected significant effects where the Robust and GLM models have indicated none, and it was found that uncentered explanatory variables, when estimating the interaction effects, can potentially introduce multicollinearity to which the Tobit models were reacting most sensitively. This provokes a general word of caution for trust researchers relying solely on Tobit models to analyze investment game data. Secondly, in addition to cross-validate the model specifications with different estimation techniques, it is advisable to re-estimate models using orthogonalized interaction variables. Overall, the combination of methods used here entails that the model of adaptive rationality can be tested without falling into the pit-trap of spurious multicollinearity and miss-specification. In combination with multiple estimation techniques and the use of bootstrapping methods to address issues of non-normality, influential data points and robust standard errors, this lends considerable credibility to the established results.

6.6. Analyzing Decision Times

6.6.1. Model Specification

The next section will detail the test of the model of adaptive rationality by examining the decision times (DT) of the subjects which were recorded during their participation in the investment game experiment. Similar to the analysis of the choice of a trusting act, this demands the specification of an empirical model to predict and test hypotheses which can be derived from the mode-selection threshold. In general, the automatic mode is expected to be fast and effortless, whereas the activation of the rational mode, paired with an increased degree of cognitive elaboration, is expected to be slow and serial. According to general model proposition 8.1, the processing modes are directly linked to the decision time of the corresponding trusting act, and main effect hypotheses for the treatment and accessibility measures have been stated in section 6.3.2 already. The predicted effect of the processing modes on DT is opposite to the predicted effects on the relative transfer decision.

Thus, a negative main effect can be predicted for *frame* and the accessibility measures *recscale* and *trustscale.* Both a cooperative framing of the trust problem and a high accessibility of trust related frames and scripts push the mode-selection threshold towards the automatic mode and lead to a relative increase of unconditional trusting strategies. This results in a *de*crease of decision times. In contrast, high initial endowments push trustors towards a controlled elaboration of the trust problem and conditional trust. Therefore, an *in*crease in DT is

expected. As with the choice of a trusting act, the model of adaptive rationality bears more complex interaction patterns among the processing mode determinants, and it encourages their analysis both theoretically and empirically. With respect to predicted interaction patterns, it is easy to show that they are opposite to those stated in chapter 6.3.3. For example, the analysis of a continuous measure of accessibility and the two experimental factors yields a set of predicted patterns in which the coefficients are reversed (table 12).

Table 12: Predicted interaction patterns for decision times (*time*)

	Predicted Interaction Pattern (Main- and Interaction Effects)																
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a_j	0	≤ 0	≤ 0	≤ 0	$= 0$	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	0
U	0	≥ 0	≥ 0	≥ 0	> 0	$= 0$	$= 0$	≥ 0	≥ 0	≥ 0	> 0	$= 0$	≥ 0	≥ 0	$= 0$	≥ 0	0
o_i	0	≤ 0	≤ 0	≤ 0	$= 0$	≤ 0	≤ 0	≤ 0	≤ 0	≤ 0	$= 0$	< 0	≤ 0	≤ 0	≤ 0	≤ 0	0
$U \cdot o_i$	0	≤ 0	< 0	≤ 0	$= 0$	$= 0$	≤ 0	≤ 0	≤ 0	≤ 0	$= 0$	$= 0$	≥ 0	> 0	$= 0$	≥ 0	0
$a_i \cdot U$	0	≤ 0	$= 0$	≤ 0	$= 0$	$= 0$	≥ 0	≥ 0	≥ 0	≥ 0	$= 0$	$= 0$	≤ 0	$= 0$	$= 0$	≥ 0	0
$a_i \cdot o_i$	0	≥ 0	$= 0$	≤ 0	$= 0$	> 0	≥ 0	≥ 0	≥ 0	≤ 0	$= 0$	$= 0$	≤ 0	$= 0$	< 0	≤ 0	0
$a_i \cdot o_i \cdot U$	0	≥ 0	$= 0$	< 0	$= 0$	$= 0$	< 0	< 0	< 0	< 0	$= 0$	$= 0$	< 0	$= 0$	$= 0$	> 0	0

Note: The table presents predicted interaction patterns between chronic script accessibility a_j, situational cues o_i and motivation U, predicting the observed DT in the investment game.

Concerning the specification of an empirical model, this corresponds to model specifications (1) and (2), as presented in section 6.5.1; specifications (3)–(5) can be straightforwardly adapted. But before estimating the models, it is advisable to take a look at the empirical distribution of DT in the sample in order to assess whether and which statistical method can be used for their analysis. The following section presents a descriptive approach to the sample and ends with a discussion of the methods that will be used.

6.6.2. Distribution of DT and Non-Parametric Analyses

Empirically, the observed decision times have a high variance (M=17.95, SD=17.84). The distribution of *time* is profoundly non-normal (skewness=5.35, kurtosis= 48.57, Skewness/Kurtosis test for normality $\chi^2(2, N=298) = 59.96$, $p<0.001$) and it includes outliers with an extremely long latency. In fact, this is a typical DT data pattern. The next table reports the percentiles of *time*, that is, the absolute DT value below which a certain percent of observations fall, and it presents the empirical values of all observations which fall outside of a two standard deviation interval above the mean (table 13):

Table 13: Percentiles of *time*, calculated from the total sample of N=298 observations

Percentile	Percentile value of *time* (seconds)
25%	8.89
50%	13.2
75%	20.27
90%	35.89
95%	44.38
> 95% (single observation values listed)	45.64, 47.43, 45.64, 47.34, 48.38 49.53, 51.61
	---- N=9 cases above 2 * SD threshold (=53.65) ----
	55.51, 58.75, 60.17, 60.34, 75.95, 78.23, 103.84, 105.56, 207.81

While most trustors decided about the choice of a trusting act in well below a minute (total sample median = 13.2s), some observations clearly fall outside of the average range, the longest observation at 207.81 seconds. Scrutinizing the N=9 extreme outliers, there is no clear relation to *reltrust*: choices include zero and full trust, and *reltrust* is relatively evenly distributed, resulting in a mean near the 50 percent mark. However, it is apparent that their inclusion can have profound effects on the parameter estimates (statistical models will therefore be recalculated in- and excluding outlier observations).

The following table cross-tabulates the conditional medians of *time* across experimental conditions. As can be seen from table 14, a shift in median DT is visible for the framing treatment, with overall shorter DT in the cooperative context. Likewise, the manipulation of the initial endowments increases median DT in the high incentive condition.

Table 14: Conditional median of *time* (s) within experimental treatment groups, N=298

Treatment / Condition		Context	
		Neutral	Cooperative
Incentives	Low	13.69	11.16
	High	16.63	12.85

In order to detail the picture of tendencies for different subgroups of the sample, the next table reports the observed conditional mean of *time* across the various groups, as well as Wilcoxon rank-sum tests for each comparison (means were computed to allow for the WRT). In the case of continuous variables, a median-split was conducted (table 15):

Table 15: Conditional mean of *time* within subgroups, N=298

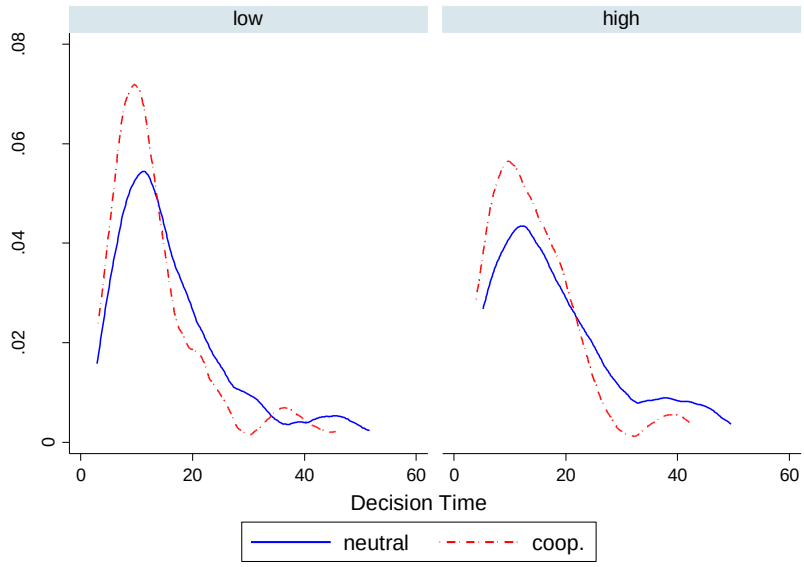
Variable	Conditional Median of *time*		Wilcoxon rank-sum test	
			Z=	*p*=
end	Low	High		
	11.85	14.35	-1.34	0.169
frame	Neutral	Cooperative		
	14.72	11.54	3.1	0.002
trustscale	Low Trust	High Trust		
	13.28	13.11	0.12	0.91
recscale	Low Reciprocity	High Reciprocity		
	13.61	11.84	1.15	0.25
fiscale	Low FI	High FI		
	13.31	12.84	-0.15	0.88
nfcscale	Low NFC	High NFC		
	13.98	11.84	1.22	0.22

The Wilcoxon rank-sum tests uncover a significant decrease in *time* for the cooperative framing condition (two sided WRT, Z=3.1, p=0.002)[22]. Descriptive evidence also exists for an effect of *end, recscale* and *nfcscale,* where a small shift in means can be observed, but none of the WRT reaches statistical significance. Thus, the alternative hypothesis that there are no differences in median *time* across these variables cannot be confidently rejected.

However, it is important to keep in mind that an exclusive focus on measures of location in the analysis of DT can be insufficient and misleading, as changes in the *shape* of the distribution are masked and cannot be uncovered (Heathcote et al. 1991). Moreover, if a distribution is highly skewed and includes outliers, then neither the mean nor the median are informative because they are potentially biased. As it is, this is the case in the present sample. To provide a visual assessment of the distribution of the DT, the next figure presents a non-parametric kernel density estimate of *time*, separated for each experimental condition. Outliers above two times the standard deviation from the mean of DT were excluded for presentational purposes (figure 24):

[22] The result does not change if a Bonferroni adjustment is conducted. The difference in *frame* stays significant.

Figure 24: Kernel density estimates of *time*, separated by experimental conditions



The kernel density estimates provide a direct assessment of the shape and distribution of *time*. As expected, the distribution is positively skewed, revealing a non-normal data pattern with a high peak at a relatively short median DT and an extended tail which results from the presence of longer decision times. More importantly, the graphs reveal differences between the experimental conditions: in the cooperative framing conditions, the density of *time* markedly increases in the lower range. Thus, with cooperative framing, more observations fall in into a short DT interval. This indicates a shift towards unconditional trusting strategies and shorter decision times. In contrast, high initial endowments result in a "fatter" tail of the distribution and lower peak densities; a hint to the presence of long decision times and prevalence of conditional trusting strategies. In combination, the results presented above suggest the presence of treatment main effects that are consistent with hypotheses H3 (shorter DT in the cooperative context) and H4a (longer DT with high endowments).

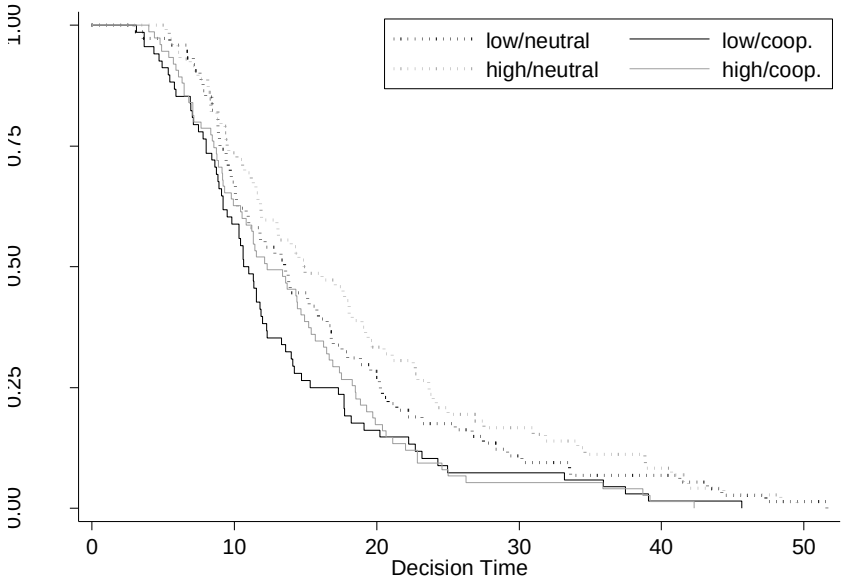
Another way to assess treatment effects is to adopt a duration-model perspective and regard the choice of a trusting act as an "event" that ends the "state" of frame- and action selection. It terminates the recorded DT interval. One can then compute and graph the conditional probability for a subject to "survive" (that is, to *not* terminate the DT measurement with a choice of a trusting act at any point in time), separated for each experimental condition. This is also known as the non-parametric Kaplan-Meier estimate of the survivor function (see Cleves et al. 2004: 93f.).[23] From the graph of the Kaplan-Meier estimates, differences in decision times

[23] In continous time, the survivor function S(t) = 1 – F(t), where F(t) is the cumulative density of the distribution. It defines the probability of survival, that is, of not observing an event conditional on its non-appearance up to time t. For discrete time intervals $t_1<t_2<\ldots t_n$, the Kaplan Meier estimator is defined as $\hat{S}(t) = \prod_{ti<t}^{t} \frac{ni-di}{ni}$, where n_i is the number of observations "at risk" at point t_i, and d_i is the number of "deaths" at t_i.

between the experimental conditions can readily be inferred (see figure 25). Holding one experimental factor constant, the level of the second factor exerts a notable influence on the estimated probability of survival. This effect is present both for the endowment and the framing conditions. Moreover, each factor affects the probability of survival in the expected direction.

High endowments lead to a higher survival probability across framing conditions. This indicates an empirical increase in DT. The effect is more pronounced in the neutral framing condition. On the other hand, a cooperative framing reduces the probability of survival. This indicates a decrease in DT. The predicted survival probabilities are most optimistic in the high/neutral condition. In this case, the probability of remaining "at risk" and observing a long DT is the highest at any point in time. In contrast, lowest probabilities can be found in the low/cooperative condition. Here, trustors have empirically made their decisions faster as in any other condition, and the probability of observing long "survivals" is the lowest across time. Adding confidence intervals around the Kaplan-Meier predictions (results omitted), a statistically significant difference in the predicted probabilities and corresponding survivor functions can be revealed between the two extreme conditions. The differences between all other subgroup comparisons are insignificant, however. Thus, while the graphical analysis indicates coherent treatment effects, their statistical effect size may be very small.

Figure 25: Kaplan-Meier probability of "failure" for the choice of a trusting act



The preceding analyses have (1) revealed the non-normal character of *time*, and (2) graphically, as well as descriptively, uncovered the presence of treatment effects, but (3) no evidence could be gathered of an influence of accessibility or processing preference measures. A simple reason may be that effects are very weak, or covered in interactive effects between treatment conditions on which main effects do not tap. To get a precise statistical estimate and test model specifications (1) to (5), it is imperative to be clear about the distributional form of DT in

the sample. The approach taken here is pragmatic: as the model of frame selection is principally open to various underlying cognitive architectures, it is not possible to derive a theoretical argument in favor of one particular response time distribution. Rather, it is advisable to use a distribution that can adequately describe the present dataset.

In a first step, using EasyFit[24], the raw measures of *time* were analyzed and fitted to a number of frequently used DT distributions (see Dolan et al. 2002, Heathcote et al. 2004), both in- and excluding the outliers of the sample. Then, goodness-of-fit measures (i.e. Kolmogorov-Smirnov tests) were computed and compared. The following table reports the results of this analysis, showing that a number of different distributions which are regularly used in response time analysis can in fact be fitted to the data (table 16):

Table 16: Fitting different distributions to the DT sample

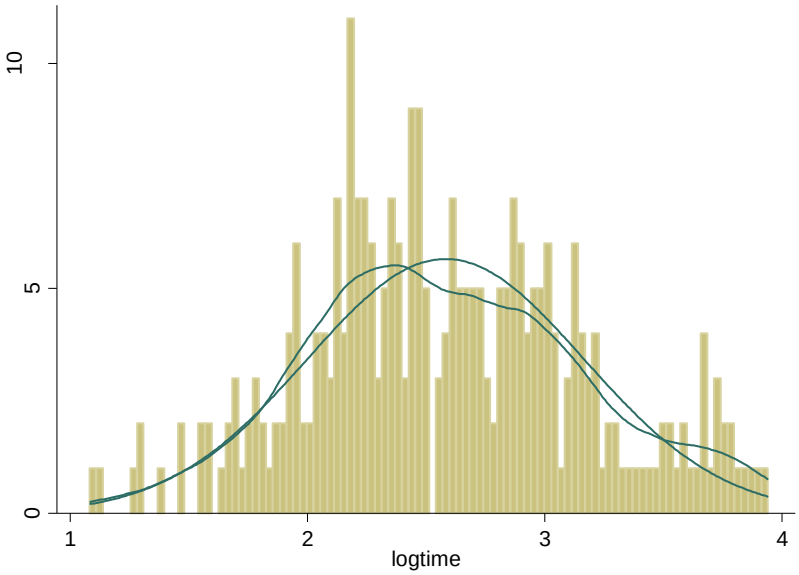
Distribution	Goodness of Fit (Kolmogorov Smirnov Test)			
	outliers excluded, N=289		outliers included, N=298	
	D=	p=	D=	p=
Lognormal	0.054	0.359	0.064	0.168
Gamma	0.067	0.135	0.210	<0.001***
Exponential	0.214	<0.001***	0.245	<0.001***
Inv. Gauss (Wald)	0.048	0.501	0.145	<0.001***
Weibull	0.113	<0.001***	0.121	<0.001***

Note: Kolmogorov-Smirnov´s D statistics report and test the maximum distance D between the assumed and the empirical cumulative distribution function. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Overall, the lognormal distribution delivers a good description of the data; most other distributions fail to adequately describe the sample with the inclusion of outliers (note, though, that even the lognormal distribution is close to being rejected with N=298). As can be seen from the table, the overall fit of any distribution decreases when outliers are included. This suggests a separate consideration of models including and excluding the extreme observations. Based on the distributional analysis, the set of model specifications will be estimated using robust linear regressions of the logarithm of *time* (*logtime*, M=2.64, SD=0.66, a histogram is presented in figure 26) on the predictors.

[24] The software is available from Mathwave in academic license at www.mathwave.com.

Figure 26: Frequency histogram of *logtime* (outliers excluded, N=289)



To control for an individual baseline speed, the variable *timeavg* is constructed as the respondent´s empirical average of two latency measures that were collected in the course of the experiment: one measure is the respondent´s decision time for a trial decision which was presented before asking the control questions. Participants had to make this trial to get acquainted with the interface. The second latency measure is the reciprocity decision in the second stage of the experiment, where subjects had to reciprocate a matched trusting choice. Here, the decision interface was nearly identical in design. Therefore this measure can serve as another approximation of the decision-making context of the choice of the trusting act, and help to pin down an individual baseline speed of response. Arguably, other factors may influence both measures over and above an individual baseline. However, out of all latency measures collected, these two measures were the only ones that could be collected in a situation that is comparable to the actual choice situation. An implicit demand of the baseline speed correction procedure is that filler latencies "match" to the target latency. In principle, one could use other latencies as well (for example, the time to answer control questions, read instructions etc.). Theoretically, they do not deliver a proper and valid baseline that can be used in the decision-making context. Empirically, it turns out that these measures are only weakly related to the actual DT measure of *time.*

6.6.3. Chronic Frame and Script Accessibility

Using the robust regression approach (see section 6.5.1), specifications (1) and (2) were estimated in order to assess the influence of chronic frame or script accessibility and processing preferences on the observed DT. Orthogonalized interaction terms were used in all models to estimate the predicted interactions between experimental treatments and the accessibility of

the trust-related frames or scripts; included controls were similar to those used in section 6.5 (see table 17):

Table 17: Regression of **c**hronic frame and script accessibility on *logtime*

Variable	Model Specification (1) Script Accessibility			Variable	Model Specification (2) Frame Accessibility		
end	0.096	0.094	0.082	*end*	0.077	0.076	0.067
	(-1.33)	(-1.22)	(-1.12)		(-0.96)	(-0.92)	(-0.85)
frame	-0.169**	-0.163**	-0.173**	*frame*	-0.165**	-0.163*	-0.168**
	(-2.37)	(-2.03)	(-2.27)		(-2.11)	(-1.90)	(-2.06)
recscale	-0.179	-0.263	-0.288	*trustscale*	0.080	0.031	-0.006
	(-0.46)	(-0.61)	(-0.71)		(-0.23)	(-0.08)	(-0.02)
*end*frame*	-2.101*	-2.050+	-1.338	*end*frame*	-0.014	0.218	0.246
	(-1.76)	(-1.61)	(-1.27)		(-0.02)	(-0.27)	(-0.32)
*end*rec*	-1.246	-1.151	-1.112	*end*trust*	0.597	0.669	0.694
	(-1.02)	(-0.89)	(-0.99)		(-0.64)	(-0.7)	(-0.76)
*frame*rec*	-1.072	-0.991	-0.448	*frame*trust*	-0.603	-0.417	-0.452
	(-0.86)	(-0.77)	(-0.43)		(-0.70)	(-0.47)	(-0.52)
*end*frame*rec*	3.024*	2.995+	2.063	*end*frame*trust*	0.004	-0.329	-0.273
	(-1.74)	(-1.63)	(-1.32)		(0)	(-0.25)	(-0.22)
timeavg	0.018***	0.018***	0.014***	*timeavg*	0.017***	0.017***	0.013***
	(-4.82)	(-4.63)	(-4.25)		(-4.4)	(-4.15)	(-4.01)
trustscale		-0.006	-0.040	*recscale*		-0.351	-0.335
		(-0.02)	(-0.12)			(-0.78)	(-0.80)
constant	2.376***	2.396***	2.507***	*constant*	2.245***	2.450***	2.551***
	(-8.57)	(-4.63)	(-5.16)		(-9.51)	(-4.76)	(-5.25)
R^2	0.156	0.113	0.149	R^2	0.137	0.151	0.130
Wald (full model)	31.92***	37.17***	37,42***	Wald (full model)	34.76***	38.91***	41.01***
χ^2 Improve (4 df)	4.01	3.63	2.2	χ^2 Improve. (4 df)	1.72	1.2	1.48
Control variables	No	Yes	Yes	Control variables	No	Yes	Yes
Ouliers included	Yes	Yes	No	Control variables	Yes	Yes	Yes

Note: N=289 excluding outliers, N=298 including outliers. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

One effect that can be reliably reproduced in all statistical estimates is a significant negative effect of the framing condition on DT. The variable *frame* reduces *logtime* about a third of its standard deviation in magnitude, indicating an overall decrease in DT and a shift to unconditional trusting strategies in a cooperative context. In contrast, a positive main effect of *end*, while it is weakly evident from the data, cannot be reliably detected. None of the control variables has a significant effect. Browsing the results, one general conclusion that can be drawn is that, even when controlling for the respondent´s baselines speed, the estimates involve a fair amount of statistical uncertainty and do not allow definite conclusions about the interaction patterns. It is noteworthy, however, that the t-values of the coefficients in model specification (1) are not "completely off" and definitely indicate the presence of interactive effects. Yet, they cannot be estimated too reliably, and Wald tests of joint significance cannot reject that the interactions are zero. An alternative specification in which the outliers were capped to the maximum of two standard deviations above the mean (see Ratcliff 1993, results omitted) pro-

duces results almost identical to column 2. Even when the outlier analysis points to a remaining influence of extreme cases, there is no *a priori* theoretical justification for their exclusion. What is more, the robust regression techniques which were used in the analysis directly accommodate for their leverage.[25]

Focusing on chronic script accessibility, the empirical signs of the interaction pattern match with predicted pattern number two. This result is remarkable because it is in line with the findings of section 6.5.2. That is to say, the model´s predictions for both the decision to trust and corresponding decision time are consistent and merge into a coherent picture. However, most t-values do not reach traditional thresholds of significance. This presents a potential type-2-error problem: should we conclude from the estimates of model specification (1) that interactive effects do not exist? It is argued here that the direct correspondence between DT and trusting choices and their combination into a consistent pattern over the domain of two dependent variables rather points to a lack of statistical power in the decision time analysis.

The following graph (see figure 27) visualizes model specification (1).[26] It presents an exploratory perspective on the model without a claim of confirmed effects. As pointed out, the interaction pattern from the regression on *logtime* mirrors the results that were uncovered in section 6.5.2. In the graph, a negative effect of *recscale* can be observed in the high/neutral, the low/neutral, and in the low/cooperative conditions. It indicates shorter DT and a shift to unconditional trusting strategies with higher script internalization. This is particularly pronounced when the "stakes are high." *Vice versa*, this finding indicates a stronger effect of high initial endowments for low accessibility subjects, who respond with longer DT and shift to conditional trusting strategies. Overall, the data suggest that high chronic script accessibility supports unconditional trust and suppresses incentive effects. This is precisely what is predicted by the model of adaptive rationality.

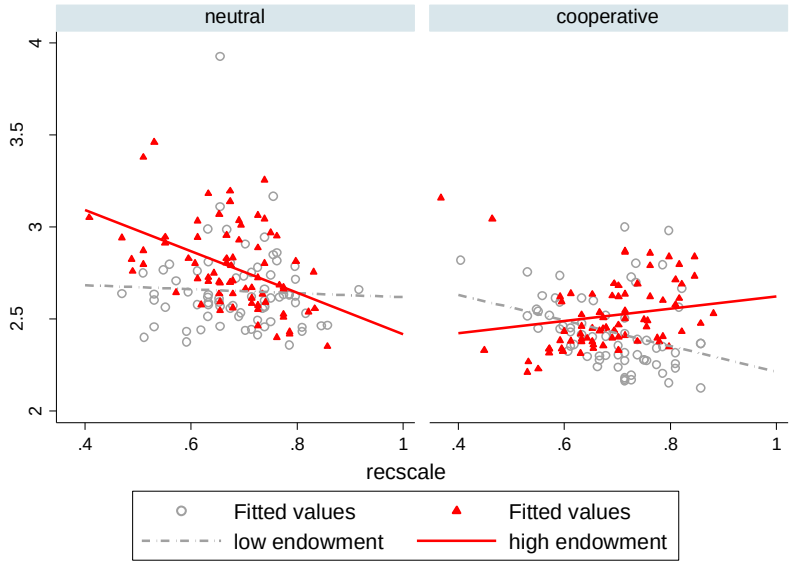
Moreover, when a cooperative frame and high initial endowments combine, the effect of *recscale* reverses in sign; the slope of *recscale* is then positive. In this case, high accessibility subjects take longer in deciding about the choice of a trusting act. In other words, the DT data weakly support the mismatch hypothesis, indicating that high reciprocity subjects have switched to conditional trusting strategies in the high/cooperative condition. High endowments may have presented a situational nuisance to them. In fact, when fixing the cooperative framing condition, a direct test reveals a weakly significant difference in *logtime* between en-

[25] As described in section 7.5.1, the robust estimation technique uses *Cook´s D* measures to generate case-wise regression weights until a stable estimate converges in which individual leverage is minimized. In the estimates of column 1, the lowest weights of the sample were in fact attached to the outliers, ranging between [0.36 0.60]. The weighting is considerably less severe for most other cases: 75 per cent of all observations were attached weights higher than 0.9; 50 per cent of the sample received weights higher than .96; 25 per cent of all cases were attached weights higher than .99.

[26] Because of insufficient statistical certainty, confidence intervals were not added to the graph and predicted values added. It was computed from the first column of table 18 including outliers.

dowment conditions for high accessibility subjects (one-sided t-test, M_{low}=2.42, M_{high}=2.58, t(2, 74)=-1.09, p=0.13).

Figure 27: Predicted values of *logtime*, using model specification (1); N=298



Another important result is that the estimate of model specification (2) in which frame accessibility is varied along with the experimental factors (presented in the right columns of table 17) is estimated to have no effects. This finding also merges with a result of the previous section. Chronic frame accessibility (as measured in terms of a generalized attitude by the "Interpersonal Trust Inventory", Kassebaum 2004) is not substantially related to the DT of a trusting act in the present experiment. Presumably, this is so because we observe a choice and not interpretation, and the reciprocity norm is a highly regulative script that is much more important in determining both processing modes and final choices than a general trust frame (see sections 6.6.5 and 6.8 for further discussion).

Again, these results must stay rather exploratory in nature. The level of noise in the DT measures is very high, and a more robust analysis would have to be built on a much larger sample. Even then, the fact that the estimated interaction patterns correspond to and match with the findings of section 6.5.2 are encouraging, and indicate that the model of adaptive rationality has a potential to predict multiple outcome measures consistently. On top of that, processing preferences might play an important role and introduce further heterogeneity, which is not captured in the current model specifications (but see below).

6.6.4. NFC/FI and Decision Times

The next analysis focuses on processing preferences and their interplay with experimental factors in the determination of *logtime*. On a general level, tensions between "intuitive" and "rational" approaches to the explanation of the trust phenomenon are an ever-present facet of

theorizing in trust research. Thus, the present experiment can inform trust researchers and substantiate this dual notion by answering the question whether, how, and when processing preferences influence the choice of a trusting act. Within the model of adaptive rationality, both NFC and FI have been introduced as additional determinants of the mode-selection threshold. Even when they have not been found to be directly related to the choice of a trusting act, an equally important question is whether trustors differ in how the information that "comes to mind" is dealt with. In this regard, processing preferences might play an important role (see table 18, all models use the robust regression approach and orthogonalized interaction variables):

Table 18: Regression of processing preferences on *logtime*

Variable	Model Specification (4)			Variable	Model Specification (5)		
end	0.069	0.070	0.069	*end*	0.102	0.099	0.088
	(-0.84)	(-0.79)	(-0.84)		(-1.39)	(-1.28)	(-1.22)
frame	-0.211***	-0.201**	-0.198**	*frame*	-0.180**	-0.172**	-0.178**
	(-2.68)	(-2.38)	(-2.50)		(-2.51)	(-2.17)	(-2.40)
fiscale	-0.267	-0.252	0.081	*nfcscale*	0.083	0.051	-0.133
	(-0.61)	(-0.56)	(-0.2)		(-0.25)	(-0.14)	(-0.39)
*end*frame*	-2.149**	-2.202**	-1.398*	*end*frame*	1.067	1.204	1.131
	(-2.42)	(-2.28)	(-1.65)		(-1.18)	(-1.3)	(-1.26)
*end*fiscale*	-1.863**	-1.898**	-1.145+	*end*nfcscale*	0.968	1.046	1.145
	(-2.21)	(-2.16)	(-1.51)		(-1.06)	(-1.13)	(-1.3)
*frame*fiscale*	-1.371+	-1.409+	-0.857	*frame*nfcscale*	-0.0142	0.059	0.014
	(-1.51)	(-1.44)	(-0.96)		(-0.02)	(-0.08)	(-0.02)
*end*frame*fi.*	3.226**	3.310**	2.192*	*end*frame*nfc.*	-1.468	-1.596	-1.41
	(-2.57)	(-2.42)	(-1.84)		(-1.23)	(-1.31)	(-1.21)
timeavg	0.018***	0.018***	0.014***	*timeavg*	0.017***	0.017***	0.013***
	(-5.19)	(-4.76)	(-4.48)		(-4.49)	(-4.27)	(-4.14)
nfcscale		-0.0673	-0.254	*fiscale*		0.287	0.438
		(-0.21)	(-0.87)			(-0.82)	(-1.43)
constant	2.467***	2.903***	2.879***	*constant*	2.225***	2.371***	2.505***
	(-7.79)	(-5.16)	(-5.47)		(-8.07)	(-4.58)	(-5.06)
R^2	0.173	0.179	0.136	R^2	0.144	0.159	0.136
Wald (full model)	47.82***	46.54***	45.83***	Wald (full model)	36.92***	39.36***	39.32***
χ^2 Improve (4 df)	7.36+	6.7+	4.4	χ^2 Improve. (4 df)	2.46	2.45	2.64
Control variables	No	Yes	Yes	Control variables	No	Yes	Yes
Ouliers included	Yes	Yes	No	Control variables	Yes	Yes	No

Note: N=289 excluding outliers, N=298 including outliers. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

In short, while there is no reliably detectable effect of NFC on *logtime* (model specification 5), the FI model delivers robust and model-consistent effects (model specification 4). This suggests that the choice of a trusting act is a matter of "feeling" and "intuition," more so than a matter of rational elaboration and "thinking." However, judging from the t-values, *some* effects of NFC can be suspected as well. At this point, it is important to remember that the NFC variable was measured with a potential ceiling effect, the empirical mean being well above the

center of the scale, with most subjects ranging high in NFC. This may have reduced statistical power and reduced the detected effect size. Even so, ceiling effects in independent variables normally *inflate* standard errors, and result in a overestimation of effects (Austin & Brunner 2003). In the current analysis, full NFC model will not be interpreted. An analysis of subgroups in the next section will show, however, that NFC can be important to some trustors as well.

The interaction pattern that is revealed in model specification (4) testing the FI processing preference is almost identical to the pattern obtained from model specification (1), in which the effect of *recscale* on DT was analyzed (a graph of predicted DT is presented in Appendix A). Since *fiscale* has a similar effect on the mode-selection threshold as the other activation weight determinants, this is in line with the set of predicted patterns that can be generated for model specification (4). One might suspect that the observed effects and interaction patterns arise from collinearity, but the empirical correlation between *fiscale* and *recscale* is weak and insignificant (rho=0.084, t=1.45, p=0.147). Furthermore, adding the control variables in column 2, which includes both recscale and trustscale, does not change the result. However, a test of the joint contribution of the interaction coefficients cannot reliably reject that they are different from zero (for example, $\chi 2$ (4) =7.36, p=0.12 in column 1). Likewise, the exclusion of outliers attenuates the result and the coefficients lose statistical precision. This raises a general concern of how to deal with outliers: while the observations are "extreme" in a statistical sense, there is no theoretical justification for their exclusion. What is more, the regression models used here are robust (down-weighing influential cases), their bias has already been taken care of.[27] As such, preference should be given to the full models.

It is important to note that a statistically significant difference in DT arises *only* in the high / cooperative condition, where high FI subjects are found to take *longer* than low FI subjects (two sided t-test, M_{low}=2.35 M_{high}=2.64, t(2, 77)=-2.73, p=0.008; all other tests are insignificant and omitted here). While we would expect a negative main effect of FI on decision times in general, the effect is positive in this case. The increase in DT in the high/cooperative condition is consistent with the proposition of a "mismatch" in this factorial constellation. The current findings suggest that highly intuitive subjects have experienced a mismatch independent of the degree of accessibility. However, a closer look reveals that there is in fact a difference, depending on whether the high FI subjects are simultaneously high in script accessibility, or not: while the positive initial endowment-effect on DT is very pronounced in high reciprocity/high FI subjects (two-sided t-test, M_{low}=2.28 M_{high}=2.65, t(2, 36)=-2.21, p=0.033), the same

[27] As with chronic script accessibility, the lowest regression weights in model specification (4) are attached to the outliers. The lowest attached weight is about 0.32 for the observation that exceeds a response time of 200 seconds; outlier weights range between [0.32 0.51]. 75 per cent of the sample were attached weights higher than .88; 50 per cent of the sample were attached weights higher than .96; and 25 per cent were attached weights higher than .99.

effect is weaker in low reciprocity/high FI subjects (two-sided t-test, M_{low}=2.43 M_{high}=2.64, t(2, 36)=--151, p=0.14).[28] This finding suggests that *processing preferences and chronic accessibility interact*, and it recommends the analysis of models in which both measures vary simultaneously. The current model specifications have not taken care of this form of interaction, and the results just presented indicate that certain subgroups of the sample will be more sensitive to the treatments than other groups. To this end, section 6.7 will continue with a more detailed exploration of sample subgroups.

6.6.5. Discussion

The analysis of decision times, as presented in the above sections, supplements the general results of the experiment in several important ways. First, the data reveal a *consistent* pattern for the influence of chronic script accessibility and experimental factors on the processing modes. The model of trust and adaptive rationality, as put forward in this work, thus receives empirical support in a multi-measure framework where predictions are generated not only for decisions, but also for corresponding decision times. This also exemplifies how the mode-selection threshold can be used to predict the emergence of different types of trust and a set of interaction patterns which can be tested against the data. Importantly, the predicted types of trusting strategies and their occurrence, that is, conditional and unconditional trust, can be compared and traced back not only to behavioral measures of trust, but also to empirical correlates of the processing mode, as measured in the form of decision time latencies.

Furthermore, and in line with the above analyses of the choice of a trusting act, the current models indicate that important determinants of trustor behavior cannot be uncovered with an analysis of simple main effects. Subgroups of the sample are heterogeneous in their response to the experimental treatments, differing on such dimensions as chronic accessibility and processing preferences. This finding is important for trust research because these variables have only been regarded in their main effect influence so far, if they have been taken care of at all. The experiment reveals that chronic accessibility and NFC/FI subgroups respond differently to the trust problem and to the experimental treatments. It is therefore imperative to accommodate for this form of heterogeneity because it can easily mask important effects. For example, a number of studies revolving around the question of "stake size" effects have uncovered inconsistent results, often finding no main effects of the manipulation, but displaying changes in the variability of the data (see Camerer & Hogarth 1999). The present experiment suggests that chronic frame and script accessibility (that is, of situationally relevant knowledge structures) may be a very important mediator of "stake size" effects. Unfortunately, none of the studies conducted in the context of trust and reciprocity have properly operationalized and

[28] Note that a Bonferroni correction would not change the results.

measured relevant frames and scripts, let alone take care of them in statistical models so far. From the adaptive rationality perspective, it is not surprising that the empirical evidence for stake size effects is weak: the models tested so far are simply miss-specified.

In the case of the collected latency measures, it is noteworthy that weak effects and a high level of noise uncovered are typical for DT data (e.g. Fazio 1986). In contrast to the experimental conditions, the participating subjects cannot be perfectly "controlled", and a number of reasons can lead to an observed latency that is well above the empirical average. Even when subjects take a long time, there is ultimately no justification for an exclusion of these observations. It is notable that the DT models provide a consistent picture from an adaptive rationality perspective. That is, estimated interaction patterns (1) merge on different dependent measures and (2) are consistent with the predictions from the theoretical model. The data presented in this work can be fruitfully used to estimate effect sizes and plan contingent follow-up experiments which directly tackle the drawback of the present study: a relatively low number of observations. As a direct methodological consequence, and since a full model test demands a higher number of cases (Kroneberg 2011c), the models will further be tested *partially* for subgroups in the remainder of the empirical part. Thus, in the light of the previous decision time analyses and findings, the next section will present more specific tests and separate hypotheses which can be generated for particular subgroups.

Overall, the data support a perspective of trust and adaptive rationality in which the question of mode-selection assumes a central role in the emergence of different types of trust. The idea of a contingent and flexible use of conditional and unconditional trusting strategies and the influence of "situated cognition" in a particular trust problem are linked to more fundamental determinants of the processing modes, such as chronic accessibility, situational cues, and cognitive motivation. The experimental manipulation of these parameters indicates that the interplay between the different mode determinants is profound and considerable, and shaping both interpretation and choice in a trust problem.

6.7. Exploring Subgroups

6.7.1. Low and High Accessibility

The findings of the multivariate analysis of *reltrust* and *logtime* suggest that there is more variability in the data than can be uncovered by restricting the analysis on simple main effects. In fact, the analysis of population subgroups is of immediate concern within the model of trust and adaptive rationality because (1) interaction effects between the mode-selection determinants are predicted and (2) concurrent hypotheses can be formulated for specific subgroups (for example, high-accessibility *versus* low-accessibility subjects). The following descriptive

analyses contrast extreme subsets of the sample, splitting observations along a number of variables of interest.

A natural candidate for a more detailed extreme-group analysis is the chronic accessibility of trust-related frames and scripts. For example, it is instructive to look at subgroups of the sample which score high on *both* measures of chronic accessibility and compare them to the low-score counterparts. Specific hypotheses can be generated for these two extremes. On the one hand, the effect of chronically accessible knowledge on trust should be strongest for the high frame / high script accessibility trustors, where unconditional trust, speaking in terms of a main effect, is most probable. With respect to empirical measurement, this translates into a high predicted *reltrust* and low predicted *logtime* measure. In contrast, low frame / low script accessibility subjects should be particularly prone to switching to elaborate processing strategies and conditional trust, with opposite and contrasting implications for trust and decision times. To conduct the following analysis, the sample was split along the corresponding median values of frame and script accessibility to identify these subgroups. The high frame / high script accessibility group includes N=63 observations, the low frame / low script accessibility group consists of N=67 observations. Splitting these groups along the experimental conditions further reduces the number of observations per cell. On average, there are N=16 observations in each cell of the next table. Therefore, while a coherent picture of tendencies emerges, these do provide moderate statistical certainty and may be subject to random sample fluctuation (table 19):

Table 19: Conditional mean of *reltrust* and *logtime* for accessibility subgroups

Treatment			Context	
			Neutral	Cooperative
Incentives	Low (7€)	*reltrust*	0.56 (0.44)	0.51 (0.46)
		logtime	2.46 (2.84)	2.19 (2.48)
	High (40€)	*reltrust*	0.60 (0.29)	0.39 (0.31)
		logtime	2.75 (2.81)	2.71 (2.52)

Note: The table presents the means of *reltrust* and *logtime*, conditional on experimental treatments, for the high frame / high script accessibility group. The numbers in brackets show the corresponding conditional mean value of *reltrust* and *logtime* for the low frame / low script accessibility subgroup.

A clear-cut tendency is revealed from the data. In any experimental condition, the relative transfer of the high accessibility group exceeds that of the low accessibility group. In addition, the recorded decision times are shorter in all factorial constellations but the high / cooperative condition. Here, it is in fact the high accessibility group which displays an increase in DT relative to the low endowment condition, and longer DT in comparison to the low accessibility contrasts. Given that the "mismatch" hypothesis holds, these observations are in line with the predictions that can be generated from the model of adaptive rationality for the particular subgroups. In a broad sense, the uncovered pattern exemplifies the prevalence of uncondition-

al trusting strategies and more automaticity in high accessibility subjects, who in general do not only trust to a higher degree, but also decide faster about the choice of a trusting act. In contrast, members of the low-accessibility subgroup, in lack of internalized knowledge structures, obviously turn to conditional trusting strategies, resulting in longer decision times and a decrease in the observed level of trust. In line with previous findings, the high / cooperative condition presents a special situation in that DT are longer for high accessibility subjects. This supports the "mismatch" interpretation as proposed in sections 6.5 and 6.6.

6.7.2. Cognitive Types

Several researchers in cognitive psychology have used the NFC/FI and related measures to construct "cognitive types" and explore differential effects of experimental treatments on subgroups which systematically differ in their processing preferences (Cacioppo et al. 1996, Shiloh et al. 2002, Betsch 2004). Similar to the extreme groups of accessibility, one can use the FI/NFC measures to identify "cognitive misers" and "cognitive monsters", that is, subgroups scoring high on one measure and low on the other. In the case of processing preferences, these identify the extreme groups. Subjects scoring high (or low) on *both* measures represent an "intermediate" case: even when the two preferences are considered to be independent, it is not clear which preference prevails in a given situation, and how the seemingly conflicting impulses from intuitive and rational processing preferences are internally compromised (Shiloh et al. 2002). Focusing on single measures, several authors have reported contrasting effects for high *versus* low NFC groups. For example, Smith and Levin (1996) found that low-NFC subjects are affected by the framing of choice problems, whereas high-NFC subjects were more resistant in attempts to change their behavior by situational cues. Shiloh et al. (2002) elaborated on these findings and showed that high FI / high NFC ("complementary thinkers") and low FI / low NFC subjects ("poor thinkers") are most prone to framing effects. They speculate that a clear and dominant thinking style, either intuitive or deliberative, fosters resistance to framing effects because, in contrast to non-differentiated thinking styles, the actors have "strong internal guides" (ibid. 425) and do not experience internal conflicts, in case of which the influence of contextual cues increases.

Similar to the procedure of constructing high *versus* low accessibility contrasts, the sample was split along the median of high and low NFC/FI medians to construct four cognitive types. The next table reports the differences between the high NFC / low FI ("rational") and low NFC / high FI ("intuitive") subgroups (table 20):

Table 20: Conditional means of *reltrust* and *logtime* FI/NFC subgroups

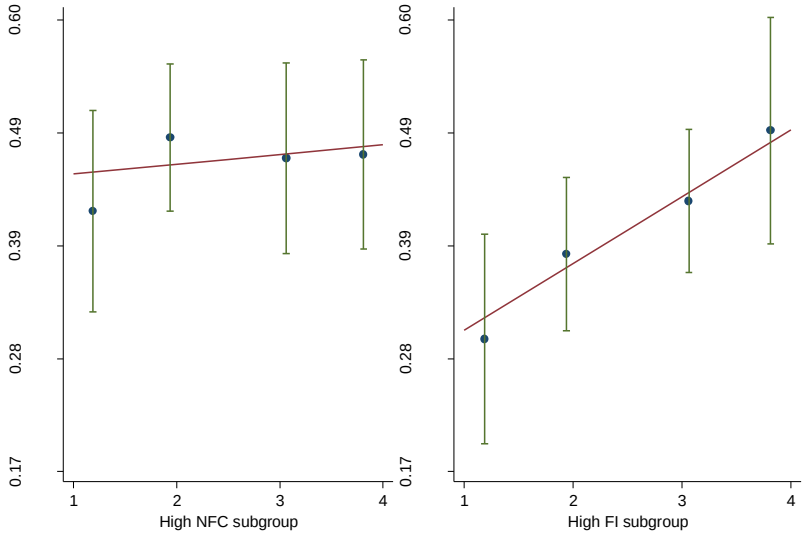
Treatment			Context	
			Neutral	Cooperative
Incentives	Low (7€)	*reltrust*	0.55 (0.37)	0.51 (0.46)
		logtime	2.48 (2.57)	2.51 (2.46)
	High (40€)	*reltrust*	0.41 (0.34)	0.34 (0.39)
		logtime	3.01 (2.69)	2.51 (2.83)

Note: The table presents the means of *reltrust* and *logtime*, conditional on experimental treatments, for the high NFC / low FI group. The numbers in brackets show the corresponding conditional mean value of *reltrust* and *logtime* for the low NFC / high FI subgroup.

A remarkable finding is that the cooperative framing condition consistently increases *reltrust* in the intuitive group across all endowment conditions, while this is not the case for the rational group. Here, a cooperative framing leads to a decrease in trust. Importantly, this replicates the findings of Shiloh et al. (2002), in that an "intuitive" but not a "rational" thinking style encourages the susceptibility to situational framing and increases the influence of peripheral cues. At the same time, the observed absolute level of *reltrust* is higher for rational as opposed to intuitive subjects in all but the high/cooperative condition. This is remarkable, as high faith in intuition should support the selection of the automatic processing mode, and therefore cater to unconditional trust. However, it is important to keep in mind that the split along FI/NFC does not inform about the chronic accessibility of trust related frames and scripts. The data therefore do not differentiate between subjects who have (or have not) available a set of trust-related frames and scripts, which potentially omits any heterogeneity along this dimension.

In fact, the average level of trust in the intuitive group is highly dependent on the chronic accessibility of trust-related knowledge, but this is not the case for the rational group. The following graph displays the conditional means of *reltrust* for the intuitive and rational subgroups, conditional on the degree of frame and script accessibility. A total of four accessibility groups were constructed by cross-splitting the observations along the medians of the corresponding accessibility measures: (1) low frame / low script accessibility, (2) low frame / high script accessibility, (3) high frame / low script accessibility, and (4) high frame / high script accessibility. These groups are used in in the following to graph the conditional means of *reltrust* for the intuitive and rational subgroups (see figure 28). Note that the graph´s x-axis represents a nominal scale; a complete rank ordering of the subgroups cannot be established.

Figure 28: Conditional means of *reltrust* for rational and intuitive subgroups by frame and script accessibility



Note: X-Axis Subgroups (1) low frame/low script (2) low frame/high script (3) high frame/low script (4) high frame/high script accessibility. Confidence intervals were constructed using confidence levels of a=0.7.

As can be seen from the graph, the intuitive subgroup, as displayed on the right, draws from the chronic accessibility of trust-related frames and scripts when deciding about the choice of a trusting act, while a similar tendency is not visible in the high NFC subgroup. The conditional means range between M_1=0.29 and M_4=0.49 for the "intuitive" trustors and increase with combined chronic accessibility. As it is, the observed average level of trust is the highest when an intuitive trustor is high in *both* accessibility measures; it is the *lowest of all subgroups* when trust-related knowledge is not accessible for an intuitive trustor. As can be seen from the added confidence intervals, neither of the differences is significant. Again, the results carry a considerable amount of statistical uncertainty and must be taken with caution. However, the analysis suggests that both accessibility and processing preferences *simultaneously* interact in determining the processing mode. This recommends an analysis of model specifications in which all parameters of the mode selection threshold vary at the same time.

6.7.3. Combining Accessibility and Processing Preferences

The goal of this this section is to explore the combination of chronic accessibility measures and processing preferences in a model in which the two continous variables and the experimental factors vary simultaneously. Put shortly, this full model includes four main effects, six two- and four three-way interactions, as well as one four-way interaction to model the mode-selection threshold. This model can be specified as:

(6) $E(reltrust|\mathbf{x}) = \mathbf{x\beta} + e = \beta_0 + \beta_1 * end + \beta_2 * frame + \beta_3 * recscale + \beta_4 * fiscale$

$+ \beta_5$*end* recscale+ β_6*end*frame+ β_7*end*fiscale + β_8*frame* recscale+ β_9*frame*fiscale + β_{10*} recscale*fiscale

$+ \beta_{11}$*end*frame*recscale + β_{12}*end*frame*fiscale+ β_{13}*end*recscale *fiscale + β_{14}*frame* recscale *fiscale

$+ \beta_{15}$*end*frame*recscale *fiscale* + controls + e

Model specification (6) displays the case for the simultaneous variation of a trust-related script and FI preferences in conjunction with the two experimental factors. A similar model can be specified using chronic accessibility of a trust-related script, by interchanging *recscale* with *trustscale,* and with respect to NFC preferences by interchanging *fiscale* with *nfcscale*. Thus, another three model specifications (7) to (9) can be estimated. A formal derivation of the predicted interaction patterns can be carried out similarly to the case with three variables. As the analysis of the current model specification is highly exploratory in nature, no predictions regarding the sign of the interactions will be made here. The principal motivation behind the analysis of the current specifications is to gauge the simultaneous influence of processing preferences and chronic accessibility in determining the processing mode. As presented in the last section, a descriptive analysis suggests that accessibility and processing preferences are not independent: high chronic script accessibility may influence the behavior of "intuitive" trustors more than that of "rational" trustors. Thus, when thinking about chronic frame and script accessibility, it is important to keep track of variable processing preferences in bringing about different types of trust. In the previous sections, they were held constant.

Model specification (6) was estimated using the familiar combination of analytic methods: first, orthogonalized interaction terms were used to ensure against spurious multicollinearity. Second, all of the different estimation methods (Tobit, Robust and GLM) and bootstrapping procedures were used to obtain robust standard errors and to account for the non-normality of *reltrust*. Third, all models were re-run with and without control variables. The findings do neither change with uncentered interactions or when including control variables. As can be seen from the statistical results (table presented in Appendix A), a number of main effects, two-way interactions, three-way interactions, and the four-way interaction are estimated to be significantly different from zero. The t-values obtained for most coefficients are considerably high and suggest that the estimated coefficients are different from zero with certainty. Moreover, the Wald tests examining the joint influence of the combined interaction terms marginally indicate that the full model improves model fit, as compared to a null model.

This result is even more striking when combining it into a broader picture with model-specifications (7) to (9). While model (7), which combines FI and chronic frame accessibility, provides relatively similar estimates of the interactive effects (results omitted), both NFC

models fail to confirm any joint contributive power of the interaction terms over a null-model, and do not result in a model that is anywhere near in statistical certainty robustness, as compared to models (6) and (7) (results omitted). This suggests that the present findings should not *per se* be ruled out under the headnote of spurious multicollinearity and be discarded as unstable. Nevertheless, the statistical test of all models that include higher-order interactions demands a high number of observations. The results are not established as confirmed effects, but remain highly explorative, presenting an outlook to the potential of the model and to future studies. The combined models reveal interactions between processing preferences, chronic accessibility and situational parameters, but the results will not be further interpreted here.

Again, it is important to keep in mind that the estimated model uses a total of N=298 observations. While it is a wide-spread practice to estimate three- and four-way interactions with much smaller sample-sizes in many (psychological) studies, the models will be taken only as evidencing *a potential for* the statistical detection of higher-order interactions between mode-selection determinants. Theoretically, such effects are implied by the model. Statistically, the present results can merely be regarded as a solid indication, and future tests would have to be built on a much larger sample.

Overall, the results of this explorative analysis are highly provocative for trust research. Principally speaking, they demonstrate that trust-related knowledge *and* processing preferences *and* situational parameters interact in determining the choice of a trusting act. This result is important because it suggests that trust and adaptive rationality are in fact much more closely intertwined than previously accepted in theory. One central idea of the current work is that we cannot think trust without thinking adaptive rationality. This statement is directly expressed in the last model specifications (6)-(9). If processing preferences shape the choice of a trusting act along with the accessibility of knowledge, then *neither* of the mode-selection determinants can be disregarded in any theoretical explanation of a choice of a trusting act.

The model of frame selection which has been put forward in this work can be used as a guide in the analysis of trust-related (experimental) data, and the empirical results uncovered in the preceding analyses provide considerable support for a perspective of trust and adaptive rationality in which mode-selections and the determinants of the processing modes acquire a central position in theorizing, model-building and causal explanation. The findings presented here can pave the way for a further empirical scrutiny of the trust and adaptive rationality perspective, and aid the development of experiments, guided by a proper theoretical model from which specific empirical hypotheses can be generated. Overall, they help to establish a broad perspective of trust that takes the aspects of interpretation and adaptive rationality to the core of its theory.

6.8. Summary of Empirical Results

The development and implementation of the empirical test in this chapter has pursued two interconnected goals: its purpose was to evaluate the adequacy of a perspective of trust and adaptive rationality in general, and to experimentally test the implications of the model of frame selection in particular. To this end, the "investment game" setting was enriched with two treatments. The trust problem was framed either neutrally or cooperative, and the stakes involved were set either to a high or a low level. These treatments allowed for a direct manipulation of two mode-selection determinants. In combination with the statistical control of the remaining parameters, this set-up provided for a direct causal test of model implications.

It was demonstrated how the model of frame selection can inform research in predicting very specific statistical effects. This does not only entail simple main effect hypotheses. The model of frame selection establishes that the mode-selection determinants interact at all stages of frame-, script-, and action-selection, and as a consequence, a number of higher-order interactions between the mode-selection determinants are predicted. This empirical specificity also attests to the high informational value and empirical content of the model of adaptive rationality. In combination with a set of bridge hypotheses that connect processing modes to observable outcomes, a set of admissible *interaction patterns* was derived against which any deviation in statistical results must be regarded as contradicting evidence.

Empirically, a number of important findings were collected. First and foremost, the framing of a trust problem and its incentive structure influence the choice of a trusting act. Thus, both experimental treatments affected the decisions of the trustors, and both exerted an influence on the corresponding decision times. While high initial endowments decrease trust and prolong decision times, a cooperative context suppresses these incentive effects and leads to a relative decrease in decision times. Both results can be interpreted as evidencing a shift in processing strategies induced by the treatments. One important mediator of treatment effects is chronic script accessibility. Thus, framing and incentive treatment effects could not be established with a simple main effects analysis only, because interactions between the mode-selection threshold parameters have to be accounted for. Most importantly, it could be shown that negative incentive effects can be fully suppressed by high chronic script accessibility. In other words, trustors who have strongly internalized a social norm may select the automatic mode and unconditional trust even when "the stakes are high." Together, these results provide direct empirical evidence of adaptive shifts in rationality, as proposed in the model of frame selection. The estimated empirical interaction pattern merges with the theoretical predictions that were generated from theoretical model.

Second, it was found that trust-related frames, as measured in the form of the "Interpersonal Trust Inventory" (Kassebaum 2004) are only weakly related to the choice of a trusting act in

the investment game. This in line with a number of other experimental results in which measures of generalized trust were found to be only weakly related to behavioral trust (see sections 3.1.3 and 6.1.1). This result pertains to models where a measure of chronic frame accessibility varies with the treatment conditions. Several potential explanations can be brought forth to understand this result. First, the measured frames may simply be irrelevant to the current experimental set-up. Even when the experiment was conducted anonymously, the trustors may have activated more specific categorical representations about their counterparts other than that of "people in general," to which most questions of the trust inventory refer (i.e. "student"). Likewise, it may be the case that the chronic accessibility of frames plays a role during interpretation, but not so during the choice of a trusting act, where the situation has already been defined. Lastly, it may be the case that trust-related frames and scripts need to be addressed in conjunction with processing preferences. If both types of mode-selection determinants are relevant to the adopted processing mode and the subsequent choice of a trusting act, then only a model which captures both effects would reveal the true relationships. In fact, the exploratory analyses in chapter 6.7 revealed a more complex interrelation between trust-related frames, scripts, and processing preferences such as "faith in intuition." In these models, trust-related frames were found to be influential in the choice of a trusting act as well. Moreover, conditional main effects for trust-related frames were also found when script accessibility was varied along with the experimental treatments.

However, the results revealed an unexpected twist in the data: the high accessibility group of trustors does not behave as expected in the high/cooperative condition. While theory would predict that unconditional trust is most probable in this subgroup/factorial combination, it was found that trustors in fact trust *less* and *increase* in their decision time. This finding was interpreted as a shift to a more rational processing and towards conditional trusting strategies; a claim that could be backed up by the analysis of decision times. Notably, this effect was not visible in low-accessibility subjects. One plausible explanation for this finding is the emergence of a selective mismatch in the particular subgroup. The situational definitions adopted by the trustors in the aftermath of a cooperative framing and the presence of high initial endowments may have created a situational nuisance and disturbed their definition of the situation. Unfortunately, this hypothesis cannot be further tested with the present data. Future experiments need to investigate the possibility of a reflexive feedback between activated interpretational knowledge structures, resulting selective attention, and the "state-dependent" attribution of meaning to situational objects.

Another important finding is that the empirically estimated interaction patterns match to the predicted patterns over the domain of two different dependent variables. This consistency in effects is a particularly powerful hint to the adequacy of the adaptive rationality perspective of trust. Concurrently, these findings suggest that an analysis of trust cannot ignore the potential

interplay between the processing mode determinants and parameters of the mode-selection threshold. A prominent example in this regard would be the question of "stake size" effects, the mixed findings of which in previous studies may be interpreted as a result of incomplete model specifications and a disregard of the potential interaction between mode-selection determinants. Adaptive rationality implies that strong norm internalizations and a high match between situational cues and accessible interpretational knowledge structures can lead to the activation of the automatic mode in which "rational" incentives may be completely suppressed, leading actors to adopting a mode of decision making in which they automatically follow their initial categorizations, activated frames, and scripts.

A result that is of interest for trust research is the finding of an influence of processing preferences on the choice of a trusting act and corresponding decision times. It was found that "faith in intuition" strongly qualifies the influence of chronic script accessibility. From a general perspective, this is not surprising: intuitive trustors should be particularly sensitive to the (non-)accessibility of trust-related knowledge. Precisely this could be observed in the data. For trust research, this is a new result that adds to our knowledge about the determinants of trust. In line with previous results from other studies, it was also found that trustors with a high "Need for Cognition" are less susceptible to framing effects and report longer decision times. This is a hint to a prevalence of more rational processing and conditional trust for the "rational" cognitive types. Both factors were also theoretically incorporated into the model of frame selection in a simple extension of the model. The practical relevance of this step must be evaluated in future studies, and it should be tied to the question of whether social groups systematically differ with respect to processing preferences (for example, academics *versus* workers), and whether these differences also translate into differential trusting behavior.

Overall, the data provide support for a perspective of trust and adaptive rationality in which contingent mode selections and a flexible degree of information processing lie at the heart of the trust phenomenon. Core propositions of the model of frame selection such as the suppressive effect of socialized frames and scripts on "rational" incentives, a flexible, dynamic and adaptive degree of rationality in interpretation and choice, and the formulation of the mode-selection threshold which determines the interplay of its parameters, can fruitfully inform trust research about the conditions that must prevail for the emergence of different types of trust.

At the same time, it is apparent that the study cannot provide the definite and ultimate answers to the modeling of adaptive rationality and the trust phenomenon. For example, the emergence of a potential "nuisance," an unexpected twist in the experimental data, brings about the question of how the match should exactly be formulated, and whether an inclusion of nuisance parameters into the match concept is obligatory or should be conducted only "on demand." The idea of "selective attention" which was put forward as a theoretical explanation here provokes the critique of being an *ad-hoc* explanative strategy and immunizing stratagem. Arguably, this

explanation cannot be directly tested with the present data, even when the results (i.e. increase in decision times connected to a decrease in the level of trust) point into a certain direction. Furthermore, in light of the results, there are two different possible causes for a switch to conditional trusting strategies: *either*, a nuisance has emerged because of a faulty study design (i.e. inadequate wording or ineffective priming procedure), *or* it has emerged because the elicited frames and scripts do in fact not extend to high-cost situations, in which case a nuisance would have emerged irrespective of the particular design features. It is impossible to discern which of the two (or a combination) has been responsible. In any case, the set of bridge hypotheses which were implicitly made in designing the experiment is incomplete. Answering these open questions (How important is "nuisance" as a relevant determinant of the mode-selection threshold? What is their cause? What is the precise domain of trust related frames and scripts such as the norm of reciprocity?) must be left to future studies.

There are further limitations of the present study. First and foremost, a common critique of experiments is their external validity, and this concern applies to the present study as well. There is a tradeoff between the power of experiments to provide an opportunity for direct testing and causal inference and the applicability of these results to the real world. This discrepancy is seen to arise from the predominant use of a homogeneous student sample, a lack of sample size and the creation of artificial situations which result in "unrealistic" data and lack practical relevance (Falk & Heckman 2009). However, as Falk and Heckman note, experiments "in the field" carry with them a different set of test conditions (for example, demographic characteristics, individual preferences, the presence or absence of social institutions and other aspects of the environment); and therefore do not automatically produce more informative results. Neither are they *per se* better suited for a test of theoretical models. In fact, experiments allow for a tight control of the conditions and constraints in which behavior takes place. This is essential for testing game theoretic models and general behavioral assumptions, as for example, the Model of Frame Selection. Therefore, experiments seem to be most powerful for the aim of testing general propositions about behavior in general, and about trust and adaptive rationality in particular. The adoption of the experimental method afforded a distinct advantage in that a direct manipulation of threshold parameters and a causal test of hypothesis was then possible. Since the model of frame selection is a general model, there is no *a priori* reason why the obtained results should be less realistic than any data gathered in the field.

Another criticism is that subjects may learn about the experiments and adjust their behavior according to the expectations of the experimenter (Hawthorne Effect). In the present experiment, the majority of participants were un-experienced first year freshmen (recruited in the second week of the curriculum) who had not participated in similar experiments before. The experiment itself was conducted as a "one-shot" game without repetition. Overall, learning and experience effects are improbable. Concerning experimenter effects, the study was con-

ducted "double-blind," that is, the experimenter was not aware of the current treatment conditions. To reduce social desirability, the participants were assured of full anonymity.

Apart from these very general points, several particular issues arise with respect to the experiment and the study of the trust phenomenon. For example, the current experiment did not control for the current emotional or mood state of the participants, even when affective influences on the choice of a trusting act have been convincingly demonstrated. Likewise, a simple investment game design was used in which neither risk aversion (Holt & Laury 2002) nor social preferences (Cox 2004, Eckel & Wilson 2004, Ashraf et al. 2006) were elicited. But in the current experiment, the randomization of subjects into treatment conditions should have equalized any systematic influence of current mood states or social preferences. What is more, the measures of trust-related frames (that is, "generalized trust") and scripts (that is, the "norm of reciprocity") which were elicited after the choice stage of the experiment can be regarded as a proxy of social preferences. As Fehr (2008) noted, social preferences are a good indicator of survey based measures of trust. If this holds, then the survey-based measures of frames and scripts have the power to accommodate and control for the influence of social preferences. Furthermore, findings about the influence of social preferences and risk aversion are relatively mixed. In a pre-test study to the current experiment, neither social preferences nor risk aversion were found to have a significant influence on the choice of a trusting act (Rompf 2008).

The current study did not seek to apply alternative measures of trust-related frames and scripts, of which a potentially endless number exists. For example, no reference was made to individual characteristics of the trustee as a basis of trust and expectation formation. Likewise, specific relational schemata, which are one of the most important types of trust-related knowledge, were only involved on the most general level of an "anonymous" counterpart who could either be a "participant" or a "partner." Presumably, these two wordings connote a different relational perspective, but they may be insufficient to activate specific relational schema. Huang and Murnighan (2010) have used a simple priming manipulation to achieve this end: the subjects had to list the names of their most favorite friends in a seemingly unrelated task. This served as a priming manipulation to increase the activation level and temporary accessibility of the specific relational schemata. While it is interesting to test these and other trust-related knowledge structures, the endeavor of an exhaustive test is not of practical relevance for a general test of the model of adaptive rationality, as long the constructs used here function properly in the mode-selection threshold. The data suggest so.

Furthermore, social norms other than the norm of reciprocity may be relevant to the choice of a trusting act and can be relevant as a regulative script. For example, fairness or equity norms may motivate trustors and the choice of a trusting act and serve as a basis for institutional trust. The present design choice was motivated by the universality of the reciprocity norm and the fact that trust and reciprocity are also structurally most intimately related (see Ostrom &

Walker 2003). Arguably, other scripts can be relevant, but this does not compromise the testability of the general theoretical model as long as the scripts elicited here do have practical relevance. Again, the data suggest so.

Concerning the question of institutional trust and how it serves as a basis for interpersonal trust, note that experiment indirectly invoked institutional trust through the framing treatment, which contained minimal references to normative institutions (i.e. the word "partner," pointing to a more communal relationship orientation with corresponding interaction norms). While no explicit institutions were designed in the experiment (no communication, no repetition, no reputation mechanisms etc.) institutional trust was highly relevant in the current experiment. In fact, the survey-based measures of generalized trust and the reciprocity norm can be interpreted as a direct approximation to a measure of relevant institutional trust. The empirical results indicate that there is a direct relation between these social institutions and trust. What is more, institutional trust can in fact become a form of "shallow trust" that is based on a rather low level of information processing and deliberation, as indicated by the DT analysis.

Next, the model test presented here was only partial in that important mode-selection determinants were not manipulated (opportunity p); others were only "held constant" or assumed to be fixed (such as the link l_i between objects and mental models). Arguably, these parameters cannot be perfectly controlled for with the present data, and the study results can be criticized along these lines: if any of the uncontrolled factors was of practical relevance in the experiment (if the participants subjectively felt time pressure, if the link between objects and mental models was a source of considerable systematic variation) *and* if these effects are not equalized and leveled out by the randomization procedure, then the current results stand on shaky grounds because we cannot allocate the variation found in the data to the independent variables introduced to the model.

Lastly, it is important to note that the total number of observations was limited with about N=300 observations. As it is, the effect sizes stemming from processes of mode-selection and adaptive rationality on a behavioral measure of trust and response times were found to be small. As a consequence, the statistical models which were estimated in the present work involve a non-negligible amount of statistical uncertainty compared to traditional benchmarks. However, more important in current model testing is whether the complete models bear explanatory power and whether the interaction variables jointly contribute to the explanation of variance in the data. It is less interesting whether a single isolated effect is significant. As was argued in chapter 6.3, the model of trust and adaptive rationality postulates interactions between the mode-selection determinants at all stages of selection. Therefore, the analysis of simple main effects can be misleading. As it is, the joint explanatory power of the interaction terms was found to be acceptable in the models analyzing trust and script accessibility, and it was acceptable in the analysis of decision times. Overall, this suggests that the uncovered ef-

fects are more than statistical artifacts and tap on substantial relations among the moderator variables. A number of results, as for example the suppression of incentive effects during the choice of a trusting act, could be established at a tolerable conventional significance level. Yet, as is always the case in any empirical study, a larger sample size would have been desirable, but it was limited by economic and logistic concerns. The present study can inform research on the question of design choice and experimentation and pave the way for future research projects to accomplish a more exhaustive and complete test of the model and all its implications.

7. Synthesis: A Broad Perspective on Trust

Trust is a subject of ongoing theoretical debate. The conceptions of trust that researchers put forward are diverse, and scholars routinely bemoan the troublesome theoretical plurality and fragmentation of trust research. A number of examples have been presented and discussed in this work. In short, the question of subjective experience is one of the prime reasons for the "confusing potpourri" (Shapiro 1987: 625) of trust definitions in the literature. Scholars focus on different phenomenological aspects and different sources of trust-related knowledge in defining the concept. As a consequence, trust definitions become too narrow and "homonymous," preventing theoretical formulations and empirical results from accumulating and becoming comparable (McKnight & Chervany 1996). Theories differ with respect to the conceptualizations, propositions, and assumptions put forward about the objective structure and subjective experience of trust. Fundamentally, they diverge on the question of how trust can be explained theoretically. Is trust a rational choice? Is it "beyond" reason or even something irrational and noncognitive? Is it an action, or a psychological state; and if so, how should this state be characterized?

This state of affairs was the impetus for the present work. As stated in the introductory chapter, a primary goal of the present thesis is to develop a broad and integrative perspective on the phenomenon of trust under a common theoretical umbrella. The guiding principle in developing this interdisciplinary perspective is to look for the commonality, mutuality, and similarities that allow the existing theory to be integrated into a broader picture; to delineate the shared theoretical and conceptual grounds on which a unifying theoretical framework for the explanation of trust can emerge. Ultimately, a broad perspective must enable scientific progress beyond descriptive work and the creation of typologies. The final destination is causal explanation, and thus a modeling of microlevel individual behavior. From the viewpoint of methodological individualism, this is the pivot around which any explanation of the social system of a trust relation must revolve. The declared purpose of the present work is to accommodate the conceptual diversity in trust research and to advance our understanding of the trust phenomenon by offering a causal reductive explanation of trust on the individual microlevel, extending it further to a macro-micro-macro explanation of trust in the spirit of methodological individualism. Notably, this broad perspective does not devaluate past research or judge one approach to be inferior to another. In contrast, it attempts to reconcile conflicting theoretical perspectives by making them understandable as a special case of a more general process.

As proposed here, two elements represent a key "missing link" to the smooth integration of the current state of the art: (1) interpretation, that is, *the subjective definition of the situation*

and (2) the actor's individual *adaptive rationality*. Both elements have been put into the focus of theorizing in this study. I argue that we can advance our understanding of trust by linking it to interpretation and adaptive rationality simultaneously. In essence, the concept of trust is fuzzy because researchers are unclear about the role of interpretation; they disagree on how trustors subjectively handle and deal with the trust problem. The process of a "definition of the situation," although crucial to the understanding of the trust phenomenon, is often mentioned in passing only, or it is taken for granted and rarely dealt with explicitly. The conversion from structure to experience, and the cognitive mechanisms involved in doing so, present a missing link in trust theory. Furthermore, I argue that *adaptive rationality constitutes a fundamental dimension of the trust concept*. There is a looming tension between cognitive and noncognitive, conditional and unconditional, rational and automatic, cognition-based and affect-based conceptions of trust that has been highlighted and emphasized throughout this book. This duality is deep-rooted and ever-present in trust research, and it has permeated to the very core of the theory, its concepts, and its definitions. But even when many authors implicitly refer to adaptive rationality when specifying the different types of trust, it has not been systematically incorporated into current theoretical frameworks, nor given the central status it deserves.

In fact, the neglect of rationality as a fundamental dimension of trust can be indeed regarded as a main barrier to the theoretical integration of existing trust research. The common ground that allows "rational" and "nonrational" accounts of trust to be united and integrated is the idea of a *dynamic, flexible, and adaptive degree of rationality* involved in interpretation and choice. This enables a seamless integration of the various typologies and approaches that have been proposed along one common and underlying dimension. Cognition-based versus affect-based, calculus-based versus identification-based, conditional versus unconditional trust: most typologies implicitly rest on specific assumptions concerning the amount of rationality involved in the choice of a trusting act. At the same time, they differentiate trust with respect to the categories of trust-related knowledge that are used by the trustor. Unfortunately, the two dimensions (category of trust-related knowledge and its "mode of application") are regularly interwoven, entangled, and regarded as fixed; the resulting typologies do not respect any flexibility in information processing when specifying the different types of trust. Essentially, current approaches do not treat adaptive rationality as a distinct dimension in its own right. But interpretation and choice, the degree of rationality involved, and the category of trust-related knowledge used to solve a trust problem are not fixed; they are independent and "orthogonal" dimensions of the typological space of trust.

But the explanation of adaptive rationality demands a focus on the process of interpretation and the subjective definition of the situation; it automatically turns our attention to the mechanisms by which the cognitive system regulates and achieves trust in "situated cognition"

(Kramer 2006). This entails a dynamic adaptation of information processing states to the current needs of the situation and the context-sensitive activation of trust-related knowledge. Methodologically, this also necessitates a clear distinction and separation from interpretation and choice. Overall, when thinking about trust, the process of the subjective definition of the situation is central in specifying the phenomenological foundations, the "mindset" of trust, and the associated subjective experiences. In order to understand trust, we must sharpen our understanding of the "missing link" of interpretation. Concurrently, it is necessary to advance our knowledge and comprehension of adaptive rationality. This cannot be done without focus on interpretation and contingent mode-selections. They jointly determine the "route to trust." The present work seeks to close this gap in current trust research.

7.1. Trust, Framing, and Adaptive Rationality

To equip trust research with the necessary tools, chapter 4 was wholly devoted to the exploration of adaptive rationality, as developed and promoted in the area of social-cognition research, in particular the dual-processing paradigm. The model of trust and adaptive rationality uses a general theory of action that directly builds on these important contributions. Using the Model of Frame Selection (MFS), I conceptualize trust as the outcome of the multi-stage process of frame, script, and action-selection. This combines separate steps of interpretation (frame and script-selection) and choice (action-selection) paired with a flexible degree of rationality at each stage in one general theoretical framework. A crucial step towards causal modeling is the capability of the MFS to bring the *determinants of information processing* into a functional relation and to spell out the *mode-selection threshold* which defines the conditions that must prevail for automatic or rational information processing to occur. Guided by the natural assessments of opportunity and motivation, and relying on the initial categorizations of unfolding pattern recognition (the activation weights and "match"), mode-selections endogenously determine the degree of rationality involved in interpretation and choice. In other words, it directs the automatic or rational selection of trust-related knowledge at each stage of the trust development process. Furthermore, the formulation of explicit *selection rules* within the MFS establishes a long needed *causal link* between cognition and action, and thus between the categories of trust-related knowledge, the processing modes, and the choice of a trusting act.

This reveals how the purported "leap of faith" and suspension in trust can be understood. As it is, suspension can occur at different stages of the trust-development process. Ultimately, it evolves from the contingent activation of the automatic mode during interpretation and choice of a trust problem. If the default mode of automatic information processing is selected and remains undisturbed, then trust can emerge without further scrutiny of the trust problem and without rising into the awareness of the trustor. The subjective experience associated with this

form of suspended, unconditional trust is nevertheless multifaceted: it may resemble the heuristic use of affect and cognitive experiences as a "quick-step," or be guided by the swift application of relational schemata, trust-related rules, roles and routines, and any other source of trust-related knowledge. What matters, in the end, is that any potential doubts or the awareness of vulnerability is suspended into subjective certainty *at the level of mode-selection*. That is, suspension is not a conscious and deliberate achievement of the trustor. It either occurs, or it does not. If conditions prevail that foster a switch to more elaborated and controlled processing of the trust-problem, then trust may ultimately acquire those characteristics which are typical of cognition-based trust, feel "bothersome," and promote a form of conditional trust in which only a "pretense" of suspension is at work. Arguably, we cannot predict which category of trust-related knowledge will come to bear in a particular solution of the trust problem. But importantly, its mode of application and the processing state of the cognitive system during interpretation and choice shape the "type" and nuance of trust that emerges in the subjective experience of the trustor as a result.

Consequentially, I have defined trust as *an actor's definition of the situation that involves the activation of mental schemata sufficient for the generation of a favorable expectation of trustworthiness and the subsequent conditional or unconditional choice of a trusting act*. This definition is very general and does not take care of the respective content of trust-related knowledge, nor demand a certain processing state. The "typological" specification of trust depends on what category of knowledge is being used, and in which mode of information processing it is applied. Nonetheless, it should be clear that any attempt of specifying a closed set of all-encompassing types of trust is futile. The definition presented here merges psychological aspects, that is, trust as a "state of mind" (or state of the cognitive system), with the behavioral aspect of choice and action. The choice of a trusting act can causally be traced back to an attempt at rational inference, at assessing trustee characteristics, and rationally weighing the expected costs and benefits of action and the activation of specific expectations, as well as to a routine execution of trust-related knowledge (relational schemata, rules, roles, routines) and a reliance on heuristic shortcuts in interpretation and choice. In the case of unconditional trust, *suspension and the "leap of faith" take place on the level of mode selection* (!), the parameters of which display the individual's history of learning and socialization. Only in conditional trust will trustors consciously access their expectation of trustworthiness. In this case, the context determines the relevance of trust-related knowledge and enables, via appropriateness beliefs, the formation and generation of expectations. I claim that this model of trust incorporates and reductively explains conditional and unconditional trust.

The model offers access to the phenomenon of interpretation and suspension; it also locates the formation of *trustworthiness expectations* in the individual framing process. The broad perspective on trust assumes, under ideal conditions, an intimate *match* between cognition and

context. Then, expectations are a direct equivalent of the *appropriateness beliefs* that pertain to the applicability of trust-related knowledge in a trust problem. There is a subtle but important difference in the conception of expectations, as put forward, for example, in the rational choice paradigm, and the present formulation as an appropriateness belief. Even when the two are practically indistinguishable in the rational mode, the framing model connects the formation of expectations to more basic cognitive processes of *schema recognition* and the activation of trust-related frames. Appropriateness beliefs point to social-psychological concepts such as "fit" and "applicability," that is, to cognitive matches between stored mental schemata and situational cues and the spreading activation occurring in response to perception. As such, they mirror the working of a basic categorization process, and an internal achievement of the cognitive system. They are *not* merely a result of knowledge retrieval, nor based on trust-related information alone. Concurrently, the context-dependent activation of frames also has the potential to explain the emergence of different social preference functions, which are treated as exogenously given in rational choice models. They become relevant only if corresponding cultural knowledge is activated *and* processed in the rational mode.

The human cognitive system directly builds on perceptual input when regulating the mode of information processing. Obviously, when taking into account human cognitive architecture, the process of trust may begin even before a conscious and deliberate interpretation of the trust problem, and without any effortful, elaborate and controlled decision-making process. This is the case when automatic interpretation and choice are furnished by salient and appropriate situational cues. If the routine of everyday behavior can be maintained by "matching" situational stimuli to preexisting stored interpretive schemes, then the allocation of attention, the conscious awareness of trust problems, and doubtful reasoning processes about the choice of a trusting act may be fully absent. One can argue with Luhmann that, in this sense, familiarity, trust and confidence do in fact gradually merge into each other.

There is an important theoretical consequence of the broad perspective on trust that I have developed here. Although it becomes possible to explain various types of trust reductively as a consequence of context-dependent framing processes and adaptive rationality, the concept of trust itself seems to dissolve and become a redundant category. Many trust researchers are concerned that trust research is in danger of becoming irrelevant, because the concept seems to refer to all and nothing at the same time. In his work on the trust concept, Möllering, for example, demands that trust research needs to claim "some unique element in the concept of trust that existing theories are not able to capture" (2006b: 9), and he identifies suspension and the "leap of faith" as these unique elements. He concludes that, "trust research needs to find out how the leap is made" (ibid. 192). If there is some substance to the conception of trust that I have offered, then trustors principally "leap" into trust during automatic mode selections. That is, suspension is a result of the very general functioning of the cognitive system, and it is

hard to claim anything unique about it. Thinking in terms of adaptive rationality, suspension is not even exclusively related to the phenomenon of trust. One unique characteristic of the trust phenomenon is the fact that the very general process is then directed towards a situation that has the structure of a trust problem, and that it is solved with the help of different categories of relevant trust-related knowledge.

In fact, I argue that there is no need to claim anything unique about trust. While it is true that, with the model of trust and adaptive rationality at hand, trust loses much of its "mysterious" and "elusive" character, it is rather a strength and advantage of a good theoretical model to make things look easy, once the hard work is done. Even though the drawing of disciplinary borders is often helpful in identifying a research domain and developing its agenda, social science, to me, is set on a route towards an integrative and interdisciplinary unification. There is no reason to exclude trust research from this development. As it is, it is one of the most interdisciplinary fields in the social sciences. It should come to no surprise that the solutions offered span disciplinary borders. The explanation offered here is very general and universal, and its reductive nature brings with it the property that a wide range of phenomena can be covered. However, it neither denies the importance of trust to social processes, nor implies that trust research is a meaningless endeavor that does not contribute to the social science agenda.

Some researchers doubt that the route towards a general approach can be taken at all. For example, Bigley and Pearce fear that "a universal conceptualization of trust and distrust may have difficulty in attaining a sufficient level of theoretical and empirical viability for research purposes" (1998: 408). That is, when "stretching" trust too far, there is a high risk of "producing constructions that are either too elaborate for theoretical purposes or relatively meaningless in the realm of empirical observation" (ibid.). In contrast, I claim that the model of trust and adaptive rationality is neither too theoretically complex, nor empirically empty. While it is true that its implications are complex and tedious to spell out, the empirical content of the theory is very high. It was derived here as a set of admissible *interaction patterns* which are implied by the model. This sort of hypothesis generation is beyond the proposition of simple main effects or the statement of general model propositions. The distinct advantage offered by the current model is that it is context-free; the relevant categories of trust-related knowledge, the frames and scripts used by trustors, are open to more detailed specification in a particular research problem. But the basic mechanism behind trust becomes transparent.

7.2. The Role of Institutions and Culture

In the model of trust and adaptive rationality, normative and cultural systems acquire a major role in the emergence and evolution of trust. Both provide and add to the stock of trust-related frames and scripts in which the social definition and constitution of a trust relation can occur.

They come, for example, in the form of social roles, norms, rules, routines, as well as cultural codes, moral standards or value systems, in sum, in the prevalent "trust culture" of a society. Together, these mental models constitute a major interactional resource on which the context-sensitive definition of a trust problem can unfold in a particular situation. This evidently implies that trust can neither be studied nor fully understood exclusively on either a purely individual or a collective level, because it thoroughly permeates both. Social institutions often create the background of familiarity on which trust becomes tangible; they also provide the structural "safeguards" and structural assurance that enables trust between individuals.

Broadly speaking, institutions and culture help to instill "taken-for-grantedness" and establish and maintain stable and unproblematic interaction. A major function of institutions is thus to provide a reduction in social complexity by providing socially shared information about the likely course of action in a social context—they do so, as proposed here, in the form of learned mental schemata about typical situations (frame), typical action sequences (scripts), typical actions by typical actors (role), and rules of action (norms). In the model of trust and adaptive rationality, these concepts are directly incorporated and mirrored in the *chronic accessibility of trust-related frames and scripts,* a crucial component of the activation weight and match. When institutions instill taken-for-granted expectations, the corresponding internalized mental schemata are often enacted without question, following a "logic of appropriateness" (March & Olsen 1989). On the individual level, this amounts to postulating a prevalence of automatic selections during interpretation and choice. When the context of a trust problem indicates that certain institutions are part of the "rules of the game," trust is enabled between actors because the institutions provide the means for a social definition of the situation and guide the individual framing processes without interruptions or nuisances. Ultimately, trust and trustworthiness can themselves acquire a taken-for-granted character in a particular and familiar trust problem. Rule-based forms of trust can trigger suspension without a conscious calculation of consequences. In specifying a causal model on the level of individual behavior, it is apparent that the mechanism behind unconditional trust is the contingent selection of the automatic mode, triggered by a high match between stored mental schemata and situational cues. I have furthermore argued that one most important class of trust-related knowledge can be found in generalized and specific relational schemata, of which humans acquire a plentitude in their social life.

Moreover, the present thesis also extends the framework of trust and adaptive rationality from the individual's to a collective, dynamic and interactive perspective. In the present conception of trust, actors normally reach the subjective definition of a trust problem in symbolic interaction with each other, relying on the dynamic process of *communication*. Essentially, any trust relation must be explained as a genetic sequence of meaningful communicative acts in which the actors' subjective definitions of the situation temporarily converge into a shared social

definition of the situation. Communication is decisive for the development of trust because it defines and influences the environment in which individual framing processes occur. This symbolic-interactionist perspective on interpersonal trust implies that *trust relations must always be reciprocally and actively defined*. Communication serves as the springboard for interpretation; it is concurrently the major vehicle for producing trust-related cues. In short, the context and environment of a trust relation cannot be treated as static. They are dynamic, and actively shaped by the involved actors, by their actions and relational communication.

The broad perspective of trust that I have developed here explains the constitution of a trust relation as a result of *reflexive social framing* (Esser 2001: 496). Social framing describes sequences of individual frame and action selections, their aggregation into a new objective social situation, and a feedback into new individual framing processes. The constitution and continuation of a trust relation then depend on structural coupling and the temporary convergence of communicated meaning. A trust relation as a social system is "locally" constituted within a particular social environment as result of social framing processes. This is guided by the application of shared frames, which are reciprocally activated during communication. But social framing processes are bound to the laws and limits of individual adaptive rationality. For unconditional trust to emerge, the chains of communication associated with a trustful course of action need to unfold without problematic interruptions, and significant symbols must be effortlessly decoded, so that a structural coupling of communicative acts smoothly accumulate into the choice of a trusting act and its trustworthy response. As mentioned before, it is a unique contribution of this work to go beyond a statement of principle relations, and to instead spell out the necessary causal conditions in a precise and tractable theoretical model.

The role of institutions and culture in this sequence cannot be underestimated. The cognitive dimension of trust, the trustor's knowledge of the social world, points to processes of learning, socialization, familiarization, generalization, and to the development of practically relevant interpretive schemes and their routine application. The ability to trust is based on past experience, learning, and familiarity with the individual life-world, which render available the different categories of trust-related knowledge: specific information, such as trustee characteristics, knowledge of dyadic and network embeddedness, and knowledge of the cultural-normative frameworks surrounding the trust relation—such as rules, roles, norms, values, relational schemata, stereotypes, and so forth.

Another important implication of the social framing perspective on trust is that a trust relation, as any social system, must always be regarded as a state of temporary balance and a fragile "quasi-stationary equilibrium." Even when trust sometimes appears as static, balanced and consistent, such an impression merely emerges from a snapshot of a dynamic time-dependent process. In this regard, the concept of *active trust* points to the flexibility and creativity in the feedback process during the social construction of trust. The actions of the parties involved

shape the emergence, continuation or dissolution of a trust relation, and trustor and trustee can influence the production of trust-related cues with relational communication, identity signaling and impression management. But even when the reflexive constitution of a trust relation is a dynamic, open and volatile process, the causal antecedents to trust reside in the psychological—and information-processing—states of individual actors. The opportunities and constraints of individual framing and bounded rationality extend to any social situation. Overall, when thinking about trust from a social framing perspective, the achievement of favorable conditions conducive to trust has to be regarded as a mutual achievement of the parties involved, and the openness and autonomy inherent in communication leaves much space for a creative element and for an opportunity to actively shape the definition of the trust problem. This opportunity relates to both the trustor and the trustee, each of whom can actively and deliberately influence the other's perspective. At the same time, it is clear that trustor and trustee rely on a large set of shared interpretive schemes during interaction. The stock of trust-related knowledge which actors use is, to a large extent, socialized and socially shared; the social construction of trust therefore always points to the cultural and institutional prerequisites of trust.

Importantly, the dynamic perspective which was added in chapter 5 also demonstrates that institutions are not just passively consumed, but actively (re-)produced in an ongoing process of symbolic interaction and reflexive structuration. They are both an objective fact of a socially constructed reality and an internalized part of individual identity at the same time (Berger & Luckmann 1966). The broad perspective promotes a symbolic-interactionist conception of trust. The constitution and social construction of trust involves the development, maintenance and application of interpretive schemes to which the actors refer, and which they symbolize and externalize during interaction. At the same time, they reproduce the social structure which is conducive to a buildup of trust and to which future action can refer. Trust, I argue here, is inseparably tied to this reflexive reproduction of structure and action. The social framing perspective of trust accommodates the idea that trust can emerge blindly in social interaction, based on routine, familiarity, and taken-for-grantedness. The present work contributes to such a structuration perspective by adding a microlevel foundation from which this process can be understood, and by delineating the role of normative and cultural systems in the cognitive processes involved.

7.3. Avenues for Future Trust Research

In this closing section, I want to reconnect my work to larger research agendas in the social sciences and highlight avenues for future research. The rise of trust as a "hot topic" is undoubtedly connected to its central role in a number of rudimentary social processes and its importance for many outcomes of human social life. An ever-increasing number of empirical

studies confirm that trust is of high social and economic relevance. A need for its theoretical explanation and for future research arises on all levels of analysis, from micro to meso and macro-analyses. Hence, the following stipulations are necessarily selective and cannot be considered exhaustive. In connecting to other research agendas, it is also apparent that cross-fertilization can always occur in both directions. Trust research draws heavily from achievements and progress made in other disciplines, and its own progress can feed back into a number of related fields, and help to shape and advance the broad research agenda of social science.

For example, trust has been regularly connected to the question of identity and the individual. It directly merges with research about the development, stability, and change in personality. It is worth noting that current notions of the "psychology of the individual" have shifted from viewing personality as a stable set of traits into a more dynamic perspective that draws heavily from the dual-processing perspective of social cognition. This view is inherent in the MFS, where identity is recast as dependent on context-sensitive activations of frames, scripts and associated schemata of the self. The concept of a relational schema, which contains schematic descriptions of both self and other in a particular context, was promoted here as a prime source of trust-related knowledge. Concurrently, relational schemata are a prime source for the adoption of individual, relational and collective identities. The social framing perspective conceptualizes identity as a dynamic and temporary state and puts the social situation and its interactive construction into the focus of interest. In connecting identity theory to the concept of social framing and adaptive rationality, psychological research is directed towards the structural, normative, and cultural antecedents of identity and "identity salience." Overall, connecting the broad perspective to the psychological and social-psychological agenda opens up a number of important avenues for future research.

For one, it is clear that the very general propositions made here about the phenomenon of trust can always be adapted to a more detailed specification in real-life social contexts. To answer the question of interpersonal trust and to explain the emergence of trust relations in an applied context requires our understanding of the "concrete" frames and scripts which become practically relevant in, for example, romantic and marital relationships, ordinate-subordinate, patient-physician, and buyer-seller relationships, and so forth. This specification could be accomplished with additional qualitative studies to help refine the measurement of trust-related frames and scripts. Moreover, a number of existent instruments for specific forms of trust could, in principle, be tested for their role as chronic frame or script accessibility indicators in the relevant contexts. This would naturally carry the empirical basis of research from experimental settings to the analysis of field data, and thus provide additional insights to the external validity of the results gathered in the present work. The qualitative specification of trust-related knowledge not only pertains to different relational contexts, it extends to intercultural

research as well. Thus, the exploration of the intercultural bases and differences in "trust culture" could be fruitfully guided by the adaptive rationality perspective. In particular, intercultural trust research has opened up an interesting debate about the generalizability of trust models (the "etic vs. emic" debate; see Dietz et al. 2010): while many trust researchers claim an *etic* position, assuming that trust concepts, models and measures are generalizable to all cultural contexts, some researchers defend an *emic* position and argue that differences in meaning and the antecedents and consequences of trust across cultural domains result in a practical non-comparability of trust models, which necessitates separate theoretical explanations. The position taken here is decidedly *etic*. While the practical specification of frames may unveil cultural differences in relevant trust-related knowledge, the adaptive rationality perspective specifies general mechanisms that are independent of the "content" of cultural-specific frames and scripts. The predictions of the MFS model thus could be tested across different cultural settings. Intercultural studies are an important area of future research by which the model of adaptive rationality and its propositions can be scrutinized.

A second important avenue for future research on the microlevel relates to the field of social cognition and the emerging field of neuroeconomics. The neuroscience of trust, still in its infancy, has the potential to become a central criterion in the evaluation of an integrative theory of trust. Simply put, if it is possible to trace back different "types" of trust to the preferential activation of different neuronal systems with the help of neuroscientific methods, then a broad theory of trust must be able to predict data on this empirical level as well. One potential source of such "hard" data comes from fRMI analyses and the study of oxytocin release in the human brain. Neuroscientific studies could support and extend the adaptive rationality perspective of trust in providing further insights and shape the solid microlevel foundation. This also directly addresses the theoretical advancement and corroboration of the MFS. Specifically, future studies need to attempt to predict the involved neuronal processes, guided by the framework of adaptive rationality and the formulation of the mode-selection threshold. At the same time, it should be clear that cognitive research will remain a most influential factor guiding the future advancement of the MFS and its theoretical formulations. A constant dialogue and transfer of knowledge between the sociological approach to a general theory of action and the field of cognitive research addressing the roots of human cognition will remain one of the most important vehicles for an advancement of the broad perspective of trust, and, concurrently, of the adaptive rationality perspective.

At the interpersonal and interorganizational level, trust is regarded as a central ingredient in explaining cooperation and as a key to understanding the development of collective action at large. As it is, trust problems are an important class of social dilemma situations. The question of their mastery is one of the most basic questions that can be asked in the social sciences. The framework of trust and adaptive rationality can support the analysis of cooperative phenome-

na in many domains. For example, organizational science, by adopting the frame-perspective of trust, could determine the scope and extent to which the internalization of trust-related knowledge influences economic outcomes within and between organizations. Is the function of trust as a "social lubricant" limited to conditions conducive to automatic mode selections? Is it possible to determine qualitatively which norms, roles, and routines are relevant parts of the "organizational culture" and the "psychological contracts" that promote trust? Such issues could be accompanied by more practical advice on how to encourage the development of unconditional trust between the cooperating actors. Of course, this line of research is not limited to the study of organizations, but it extends to all forms of cooperation. Ultimately, one can expect new insights even in more distant areas, such as research on social closure, where the cooperative effort of exclusion and monopolization rests, to a large extent, on trust among the participating actors. From a social network perspective, stipulations for future research arise on all levels of embeddedness. Trust is a defining element in many network theories. New questions emerge once adaptive rationality is taken into account. If trust rests on adaptive rationality and networks rest on trust, then what can be learned about the stability of social networks? What does the structuration perspective of trust and its social construction imply for the conceptions of trust used in network theory? Also, how do other structural parameters, such as the distribution of power and control, influence, and potentially override, the constitution of trust within social networks?

On the macrolevel, trust has been assigned a crucial role in the question of establishing and maintaining social order and social change at large. Trust is inseparably tied to the functioning and stability of social systems by its integrative function as a "lubricant" of cooperation and efficient mechanism for the reduction of social complexity. These broad research agendas belong to the core of sociological thinking and have always been at the center of theorizing in the social sciences. The broad perspective of trust emphasizes the potential for the emergence of stable "systems of trust" and the recursive and self-enforcing structuration of social systems on the basis of automatic and routine action, in which the "logic of appropriateness" unfolds under the conditions of adaptive rationality. This opens up a huge number of avenues for future trust research, both on the theoretical and empirical front. Concerning the question of institutionalization and agency, it is interesting to explore how and when trust becomes self-reinforcing. Thus, future studies would need to address the question of how trust can stabilize into "systems of trust" (Coleman 1990) and ask about the role of framing and adaptive rationality in this regard. More generally speaking, future research needs to address the "logic of aggregation" and explore the dynamic feedback process of the social constitution of larger social systems in which trust is critical. Turning to more practical considerations, the adaptive rationality perspective has a potential to fundamentally change the way in which we look at, for example, the "psychology of markets" and individual market behavior, and similarly, the functioning of the political system. The theoretical consequences of postulating a reductive

logic of action that can easily depart from rational choice considerations is profound and needs to be gauged in future work. This can also lead to more practical conclusions concerning the question of the stability of the political and economic systems, in both of which trust is seen to play a central role. An issue that was decidedly excluded from the present work as an explanandum in its own right is *institutional trust*, that is, trust towards objects that are not individual actors. The adaptive rationality perspective bears important implications for the explanation and emergence of system trust, which back up the functioning of the institutional and cultural systems of society.

Apart from these very general suggestions, there are a number of concrete next steps which need to be tackled in advancing the adaptive rationality perspective of trust. In part, they emerge as a direct consequence of the limitations of the present work and other previous studies. To begin with, this study did not vary all factors of the mode-selection threshold. Clearly, there is much room for future experiments to manipulate other factors and other factorial combinations of the mode-selection threshold parameters (for example, in combination with opportunity). This research is necessary in further testing the precise interplay of the mode-selection determinants. Coincidentally, such studies can provide corroboration and a check on the robustness of the results obtained here. Secondly, any experimental study is confronted with the potential criticism of being not externally valid. Future studies should carry the basic framework adopted here "to the field" and devise experiments in a natural setting in which the basic propositions of the MFS perspective can be thoroughly tested. Third, the present study has not sought to explore the role of affect during interpretation and choice. If emotions are not only a consequence of interpretation, but also influence processing states, then an important avenue for future studies is to explore empirically and theoretically their role in the trust development process. On a theoretical level, the impact of emotions has only been provisionally explored within the MFS framework (see Esser 2005). Fourth, it is clear that the frames and scripts used in this experiment to operationalize chronic accessibility parameters can, in principle, be substituted by other indicators. For example, fairness norms or distributional concerns might become relevant in the experimental setting. Future experiments could devise alternative measures, the selection of which should be adapted to the concrete design of the experimental context. Thus, research needs to simultaneously explore other means of operationalizing the threshold-parameters and manipulating the context of the trust problem experimentally. Fifth, one particular issue that has emerged in the present experiment is the question of "nuisance" and its theoretical inclusion in or exclusion from the activation weights. Future experiments could be devoted to exploring the role of disruptions *versus* stability in the subjective definition of the situation. This approach could be very fruitfully connected to trust research: a prominent case of a "nuisance" in pre-established trust relations is the failure of trust by the trustee. Put shortly, one could not only experimentally create a certain context, but could dynamically change the definitions of the situation of the participants.

For example, depending on "situational strength" and norm-internalization, how stable is trust against violations in repeated interactions? How likely are trustors to change their perspectives following a breach of trust? How can trust be re-established? Sixth, I propose to re-examine previous "stake-size" experiments and studies of incentive effects for their potential use under the head-note of adaptive rationality. Thus, any data-set in which a stake-size manipulation is paired with an indicative measure of a relevant norm and its accessibility can be used to test the MFS predictions in retrospective (see for example, Johansson-Stenman et al. 2005, who collect a one-item generalized trust measure, but do not test interactive effects with the incentive treatment, as implied by the MFS). In principle, cognitive motivation could also be varied with an alternative manipulation, such as "fear of invalidity" (Sanbonmatsu & Fazio 1990). However, it should be clear that the use of monetary incentives is the most credible course of action from an "economic" point of view. Future experiments could also carry the experimental setting to market economies in which a much higher "stake size" can be achieved with the research funds (of course, cultural differences in the measurement of relevant frames and scripts have to be respected). Last but not least, future studies need to address the pending issue of generating a sufficient number of observations in the data sample. This is particularly important with respect to the analysis of decision times, which naturally have a large variation. Methodically, one potential route would be a turn to other experimental designs (potentially departing from the topic of trust and the investment game setup) which provide data on multiple observations "within" individuals. A shift from between-subject to within-subject designs with repeated measurements would allow for an even more stringent analysis of causal effects.

Summarizing these suggestions for future trust research, and restating the main argument that was developed in this book, it is crucial to recognize the importance of interpretation and adaptive rationality to our understanding of the trust phenomenon. A perspective that puts human bounded rationality to the core of theorizing but goes beyond mere descriptive work has the potential to change the way we think about a number of social phenomena in which trust plays a decisive role. Conceptually, this means that our "models of man" need to be adjusted accordingly. The perspective of adaptive rationality has the potential to provide the explanative core of a macro-micro-macro model in which the causal explanation of social phenomena can be accomplished reductively, that is, by reference to a more general process in which rational *versus* irrational, cognitive *versus* noncognitive, automatic *versus* controlled "types" of action can be traced back to a common underlying mechanism. This perspective points to the interaction between several cognitively relevant parameters that guide the adjustment of the degree of rationality involved in interpretation and choice. If this is recognized, then I am confident that future trust research will bring substantive benefits to the broad research agenda of sociology, and contribute to the advancement of social science.

8. References

Aarts, H., and A. Dijksterhuis. 2000. "Habits as Knowledge Structures: Automaticity in Goal-Directed Behavior." *Journal of Personality and Social Psychology* 78(1):53-63.

Abelson, R. P. 1981. "Psychological Status of the Script Concept." *American Psychologist* 36(7):715-729.

—. 1995. "Attitude Extremity." Pp. 25-41 in *Attitude Strength: Antecedents and Consequences*, edited by R. E. Petty, and J. A. Krosnick. Mahwah, NJ: Lawrence Erlbaum Associates.

Adolphs, Ralph. 2002. "Trust in the Brain." *Nature Neuroscience* 5(3):192-193.

—. 2003. "Congitive Neuroscience of Human Social Behavior." *Nature Reviews, Neuroscience* 4:165-177.

Adolphs, Ralph, Daniel Tranel, and Antonio R. Damasio. 1998. "The Human Amygdala in Social Judgment." *Nature* 393(4):470-474.

Aiken, L. S:, and S. G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Ainsworth, Mary D. S., Mary C. Blehar, Everett Waters, and Sally Wall. 1978. *Patterns of Attachment: A Psychological Study of the Strange Situation*. New York: Erlbaum.

Ainsworth, Mary D. S., and Carolyn G. Eichberg. 1991. "Effects on Infant-Mother Attachment of Mother's Unresolved Loss of an Attachment Figure or Other Traumatic Experience." Pp. 160-183 in *Attachment Across the Life Cycle*, edited by Collin M. Parkes, Joan Stevenson-Hinde, and Peter Marris. London, New York: Routledge.

Ajzen, I. 1985. "From Intentions to Actions: A Theory of Planned Behavior." Pp. 11-39 in *Action-Control: From Cognition to Behavior*, edited by J. Kuhl, and J. Beckmann. Heidelberg: Springer.

Ajzen, I., and M. Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. London: Prentice-Hall.

Alge, B. J., C. Wiethoff, and H. J. Klein. 2003. "When does the Medium Matter? Knowledge-Building Experiences and Opportunities in Decision-Making Teams." *Organizational Behavior and Human Decision Processes* 91(1):26-37.

Allais, Maurice. 1953. "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine." *Econometrica* 21(4):503-546.

—. 1979. "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School." Pp. 27-145 in *Expected Utility Hypotheses and the Allais Paradox*, edited by Maurice Allais, and Ole Hagen. Boston, London: Dordrecht.

Allison, S., and David M. Messick. 1990. "Social Decision Heuristics in the Use of Shared Resources." *Journal of Behavioral Decision Making* 3(3):195-204.

Almond, Gabriel A., and Sydney Verba. 1972. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton, NJ: Princeton University Press.

Anderhub, Vital, Dirk Engelmann, and Werner Güth. 2002. "An Experimental Study of the Repeated Trust Game with Incomplete Information." *Journal of Economic Behavior & Organization* 48(2):197-216.

Andersen, S. M., I. Reznik, and L. M. Manzella. 1996. "Eliciting Facial Affect, Motivation, and Expectancies in Transference: Significant-Other Representations in Social Relations." *Journal of Personality and Social Psychology* 71(6):41-57.

Andersen, Susan M., and Serena Chen. 2002. "The Relational Self: An Interpersonal-Cognitive Theory." *Psychological Review* 109(4):619-645.

Anderson, John R. 1995. *Learning and Memory: An Integrated Approach*. New York: Wiley.

Anderson, Lynda A., and Robert F. Dedrick. 1990. "Development of the Trust in Physician Scale: A Measure to Assess Interpersonal Trust in Patient-Physician Relationships." *Psychological Reports* 67(3):1091-1100.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal, The* 100(401):464-477.

—. 1995. "Warm Glow versus Cold Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." *Quarterly Journal of Economics* 110(1):1-21.

Anscombe, Francis J, and Robert Aumann. 1963. "A Definition of Subjective Probability." *Annals of Mathematical Statistics* 34(1):199-205.

Arrow, Kenneth. 1974. *The Limits of Organization*. New York: Norton, York.

—. 1987. "Rationality of Self and Other in an Economic System." Pp. 201-216 in *Rational Choice: The Contrast between Economics and Psychology*, edited by Robin M. Hogarth, and Melvin W. Reder. Chicago: Chicago University Press.

Ashmore, Richard D., Kay Deaux, and Tracy McLaughlin-Volpe. 2004. "An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality." *Psychological Bulletin* 130(1):80-114.

Ashraf, Nava, Iris Bohnet, and Nikita Piankov. 2006. "Decomposing Trust and Trustworthiness." *Experimental Economics* 9(3):193-208.
Austin, Peter C., and Lawrence J. Brunner. 2003. "Type I Error Inflation in the Presence of a Ceiling Effect." *American Statistical Association* 57(2):97-104.
Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
Bacharach, Michael, and Diego Gambetta. 2001. "Trust in Signs." Pp. 148-184 in *Trust in Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Bacharach, Michael, Gerardo Guerra, and Daniel John Zizzo. 2007. "The Self-Fulfilling Property of Trust: An Experimental Study." *Theory and Decision* 63(4):349-388.
Bachmann, Reinhard. 1998. "Conclusion: Trust - Coneptual Aspects of a Complex Phenomenon." Pp. 298-322 in *Trust Within and Between Organizations*, edited by C. Lane, and R. Bachmann. Oxford: Oxford University Press.
Baier, Anette. 1986. "Trust and Antitrust." *Ethics* 96(2):231-260.
Baldwin, M. W., S. E. Carrell, and D. F. Lopez. 1990. "Priming Relationship Schemas: My Advisor and the Pope are Watching Me from the Back of My Mind." *Journal of Experimental Social Psychology* 26(5):435-454.
Baldwin, Mark W. 1992. "Relational Schemas and the Processing of Social Information." *Psychological Bulletin* 112(3):461-484.
Banaji, Mazharin R., and Deborah A. Prentice. 1994. "The Self in Social Contexts." *Annual Review of Psychology* 45(1):297-332.
Baran, Nicole M., Paola Sapienza, and Luigi Zingales. 2010. "Can We Infer Social Preferences From the Lab? Evidence From the Trust Game." in *NBER Working Paper Series, No. 15654*. National Bureau of Economic Research (NBER), Cambridge MA.
Barber, Bernard. 1983. *The Logic and Limits of Trust*. New Brunswick, NJ: Rutgers University Press.
Bargh, J. A., and K. Barndollar. 1996. "Automaticity in Action: The Unconscious as Repository of Chronic Goals and Motives." Pp. 457-481 in *The Psychology of Action: Linking Cognition and Motivation to Behavior*, edited by P. M. Gollwitzer, and J. A. Bargh. New York: Guilford Press.
Bargh, J. A., and T. L. Chartrand. 1999. "The Unbearable Automaticity of Being." *American Psychologist* 54(7):462-479.
Bargh, J. A., P. M. Gollwitzer, A. Lee-Chai, K. Barndollar, and R. Trotschel. 2001. "The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals." *Journal of Personality and Social Psychology* 81(6):1014-1027.
Bargh, J. A., and F. Pratto. 1986. "Individual Construct Accessibility and Perceptual Selection." *Journal of Experimental Social Psychology* 22(4):293-311.
Bargh, John A., Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71(2):230-244.
Bassili, J. N., and J. F. Fletcher. 1991. "Response-Time Measurement in Survey Research. A Method for CATI and a New Look at Non-Attitudes." *Public Opinion Quarterly* 55(3):331-346.
Bassili, J. N., and J. P. Roy. 1998. "On the Representation of Strong and Weak Attitudes About Policy in Memory." *Political Psychology* 21(4):107-132.
Bassili, J. N., and B. S. Scott. 1996. "Response Latency as a Signal to Question Problems in Survey Research." *Public Opinion Quarterly* 60(3):390-399.
Bassili, John N. 1996. "Meta-Judgmental Versus Operative Indexes of Psychological Attributes: The Case of Measures of Attitude Strength." *Journal of Personality and Social Psychology* 71(4):637-653.
Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. "Guilt in Games." *American Economic Review* 97(2):170-176.
—. 2009. "Dynamic Psychological Games." *Journal of Economic Theory* 144(1):1-35.
Bauer, Hans H., Marcus M. Neumann, and Anja Schüle. 2006. *Konsumentenvertrauen*. München: Verlag Franz Vahlen.
Bauernschuster, Stefan, Oliver Falck, and Niels Große. 2010. "Can Competition Spoil Reciprocity? –A Laboratory Experiment." in *CESifo Working Paper, No. 2923*. Munich Society for the Promotion of Economic Research (CESifo), Munich.
Baumeister, R. F., E. Bratslavsky, M. Muraven, and D. M. Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74(5):1252-1265.
Baumgartner, Thomas, Markus Heinrichs, Aline Vonlanthen, Urs Fischbacher, and Ernst Fehr. 2008. "Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans." *Neuron* 58(4):539-650.
Becker, Lawrence C. 1996. "Trust as Noncognitve Security about Motives." *Ethics* 107(1):43-61.
Beckert, Jens. 2006. "Trust and Markets." Pp. 319-331 in *Handbook of Trust Research*, edited by Reinhard Bachmann, and Akbar Zaheer. Cheltenham, UK; Northampton,USA: Edward Elgar Publishing.

Bellamare, C., and S. Kroeger. 2007. "On Representative Social Capital." *European Economic Review* 51(1):183-202.
Ben-Ner, Avner, and Freyr Halldorsson. 2010. "Measuring Trust: Which Measure Can Be Trusted?" *Journal of Economic Psychology* 31(1):64-79.
Ben-Ner, Avner, and Louis Putterman. 2009. "Trust, Communication and Contracts: An Experiment." *Journal of Economic Behavior & Organization* 70(1):106-121.
Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10(1):122-142.
Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality. A Treatise in the Sociology of Knowledge*. New York: Doubleday & Co.
Best, Henning, and Thorsten Kneip. 2011. "The Impact of Attitudes and Behavioral Costs on Environmental Behavior: A Natural Experiment on Household Waste Recycling." *Social Science Research* 40:917-930.
Betsch, Cornelia. 2004. "Präferenz für Intuition und Deliberation (PID): Inventar zur Erfassung von Affekt- und Kognitionsbasierten Entscheidungen." *Zeitschrift für Differentielle und Diagnostische Psychologie* 25(4):179-197.
Bicchieri, Christina. 2002. "Covenants without Swords: Group Identity, Norms, and Communication in Social Dilemmas." *Rationality and Society* 14(2):192-228.
—. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge, NY: Cambridge University Press.
Bicchieri, Christina, John Duffy, and Gil Tolle. 2004. "Trust Among Strangers." *Philosophy of Science* 71(3):286-319.
Bierhoff, Hans-Werner, and Bernd Vornefeld. 2004. "The Social Psychology of Trust with Applications in the Internet." *Analyse & Kritik* 26:48-62.
Bies, R. J., and T. M. Tripp. 1996. "Beyond Distrust: "Getting Even" and the Need for Revenge." Pp. 246-260 in *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer, and Tom R. Tyler. Thousand Oaks, CA: Sage.
Bigley, Gregory A., and Jone L. Pearce. 1998. "Straining for Shared Meaning in Organization Science: Problems of Trust and Distrust." *The Academy of Management Review* 23(3):405-421.
Bizer, George Y., and Jon A. Krosnick. 2001. "Exploring the Structure of Strength-Related Attitude Features: The Relation Between Attitude Importance and Attitude Accessibility." *Journal of Personality and Social Psychology* 81(4):566-586.
Blau, Peter M. 1964. *Exchange and Power in Social Life*. New York: Wiley.
Bless, Herbert, Gerald L. Clore, Norbert Schwarz, Verena Golisano, Christina Rabe, and Marcus Wölk. 1996. "Mood and the Use of Scripts: Does a Happy Mood Really Lead to Mindlessness?" *Journal of Personality and Social Psychology* 71(4):665-679.
Bless, Herbert, and Klaus Fiedler. 2006. "Mood and the Regulation of Information Processing and Behavior." Pp. 65-84 in *Hearts and Minds: Affective Influences on Social Cognition and Behavior*, edited by Joseph P. Forgas. New York: Psychology Press.
Bless, Herbert, Johannes Keller, and Eric R. Igou. 2009. "Metacognition." Pp. 157-178 in *Social Cognition. The Basis of Human Interaction*, edited by Fritz Strack, and Jens Förster. New York, London: Psychology Press.
Blumer, Herbert. 1969. *Symbolic Interactionism: Perspective and Method*. Berkeley: University of California Press.
—. 1974. "Society as Symbolic Interaction." Pp. 145-153 in *Symbolic Interaction*, edited by Jerome G. Manis, and Berard N. Meltzer. Boston: Allen and Bacon.
Bodenhausen, Galen V., and Kurt Hugenberg. 2009. "Attention, Perception, and Social Cognition." Pp. 1-22 in *Social Cognition. The Basis of Human Interaction*, edited by Fritz Strack, and Jens Förster. New York, London: Psychology Press.
Bohnet, Iris, and Yael Baytelman. 2007. "Institutions and Trust: Implications for Preferences, Beliefs, and Behavior." *Rationality and Society* 19(1):99-135.
Bohnet, Iris, Bruno S. Frey, and Steffen Huck. 2001. "More Order with Less Law: On Contract Enforcement, Trust, and Crowding." *American Political Science Review* 95(1):131-144.
Bohnet, Iris, and Steffen Huck. 2004. "Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change." *American Economic Review: Papers and Proceedings* 94(2):362-366.
Bohnet, Iris, Steffen Huck, Heike Harmgart, and Jean-Robert Tyran. 2005. "Learning Trust." *Journal of the European Economic Association* 3(2):322-329.
Bohnet, Iris, and Richard Zeckhauser. 2004. "Trust, Risk, and Betrayal." *Journal of Economic Behavior & Organization* 55(4):467-484.
Bolton, Gary E., E. Katok, and A. Ockenfels. 2004. "How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation." *Management Science* 50(11):1587-1602.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90(1):166-193.
Boudon, Raymond. 2003. "Beyond Rational Choice Theory." *Annual Review of Sociology* 29:1-21.
Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Cambridge, MA: Harvard University Press.
—. 1985. "The Forms of Capital." Pp. 241-258 in *Handbook of Theory and Research for the Sociology of Education*, edited by J. G. Richardson. New York: Greenwood.
Bowlby, John M. 1969. *Attachment and Loss: Vol. 1. Attachment*. New York: Basic Books.
—. 1973. *Attachment and Loss: Vol. 2. Separation, Anxiety, and Anger*. New York: Basic Books.
—. 1979. *The Making and Breaking of Affectional Bonds*. London: Tavistock.
—. 1980. *Attachment and Loss: Vol. 3. Loss, Sadness, and Depression*. New York: Basic Books.
Box-Steffensmeier, J. M., and B. S. Jones. 1997. "Time is of Essence: Event History Models in Political Science." *American Journal of Political Science* 41(4):1414-1461.
Bracht, Juergen, and Nick Feltovich. 2008. "Efficiency in the Trust Game: An Experimental Study of Precommitment." *International Journal of Game Theory* 37(1):39-72.
—. 2009. "Whatever You Say, your Reputation Precedes You: Observation and Cheap Talk in the Trust Game." *Journal of Public Economics* 93(9):1036-1044.
Braithwaite, Valerie. 1998. "Communal and Exchange Trust Norms: Their Value Base and Relevance to Institutional Trust." Pp. 46-74 in *Trust and Governance*, edited by Valerie Braithwaite, and Margaret Levi. New York: Russel Sage Foundation.
Breckler, Steven J. 1984. "Empirical Validation of Affect, Behavior and Cognition as Distinct Components of Attitude." *Journal of Personality and Social Psychology* 47(6):1191-1205.
Bretherton, Inge. 1985. "Attachment Theory: Retrospect and Prospect." *Monographs of the Society for Research in Child Development* 50(1/2):3-35.
Brewer, M. B. 1979. "In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis." *Psychological Bulletin* 86(2):307-324.
Brewer, M. B., and R. M. Kramer. 1986. "Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing." *Journal of Personality and Social Psychology* 50(3):543-549.
Brewer, Marilynn B. 1986. "Ethnocentrism and its Role in Interpersonal Trust." Pp. 345-360 in *Scientific Inquiry and the Social Sciences: A Volume in Honor of Donald T. Campbell*, edited by M. B. Brewer, and B. Collins. San-Francisco: Jossey-Bass.
—. 2008. "Depersonalized Trust and Ingroup Cooperation." Pp. 215-232 in *Rationality and Social Responsibility. Essays in Honor of Robyn Mason Dawes*, edited by Joachim Krueger. New York: Psychology Press.
Brewer, Marilynn B., and Wendi Gardner. 1996. "Who is this "We"? Levels of Collective Identity and Self Representation." *Journal of Personality and Social Psychology* 71(1):83-93.
Bromiley, Philip, and Larry L. Cummings. 1995. "Transaction Costs in Organizations with Trust." Pp. 219-47 in *Research on Negotiations in Organizations, Vol. 5*, edited by Robert J. Bies, Blair H. Sheppard, and Roy J. Lewicki. Greenwich, CT: JAI Press.
Bromiley, Philip, and Jared Harris. 2006. "Trust, Transactions Cost Economics, and Mechanisms." Pp. 124-143 in *Handbook of Trust Research*, edited by Reinhard Bachmann, and Akbar Zaheer. Cheltenham, UK; Northampton, USA: Edward Elgar Publishing.
Brown, Rupert. 2000. "Social Identity Theory: Past Achievements, Current Problems and Future Challenges." *European Journal of Social Psychology* 30(6):745-778.
Bruhin, Adrian, Helga Fehr-Duda, and Thomas F. Epper. 2010. "Risk and Rationality: Uconvering Heterogeneity in Probability Distortion." *Econometrica* 78(4):1375-1412.
Buchan, Nancy R., Rachel T. A. Croson, and Robyn M. Dawes. 2002. "Swift Neighbours and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange." *American Journal of Sociology* 108(1):168-206.
Buchan, Nancy R., Rachel T. A. Croson, and Sara Solnick. 2008. "Trust and Gender: An Examination of Behavior and Beliefs in the Investment Game." *Journal of Economic Behavior & Organization* 68(3/4):466-476.
Buchan, Nancy R., Eric J. Johnson, and Rachel T. A. Croson. 2006. "Let's Get Personal: An International Examination of the Influence of Communication, Culture and Social Distance on Other Regarding Preferences." *Experimental Economics* 60:373-398.
Bugental, D. B. 2000. "Acquisition of the Algorithms of Social Life: A Domain-Based Approach." *Psychological Bulletin* 126(2):187-219.
Burgoon, Judee K. 1993. "Interpersonal Expectations, Expectancy Violations, and Emotional Communication." *Journal of Language and Social Psychology* 12(1):30-48.
Burgoon, Judee K., Thomas Birk, and Michael Pfau. 1990. "Nonverbal Behaviors, Persuasion, Credibility." *Human Communication Research* 17(1):140-169.

Burgoon, Judee K., and Jerold I. Hale. 1984. "The Fundamental Topoi of Relational Communication." *Communication Monographs* 51(3):193-214.
—. 1987. "Validation and Measurement of the Fundamental Themes of Relational Communication." *Communication Monographs* 54(1):19-41.
Burgoon, Judee K., and Gregory D. Hobbler. 2002. "Nonverbal Signals." Pp. 240-299 in *Handbook of Interpersonal Communication*, edited by Mark L. Knapp, and John A. Daly. Thousand Oaks: Sage Publications.
Burke, Peter J., and Jan E. Stets. 1999. "Trust and Commitment through Self-Verification." *Social Psychology Quarterly* 62(4):347-366.
Burnham, Terence, Kevin McCabe, and Vernon L. Smith. 2000. "Friend-or-Foe Intentionality Priming in an Extensive Form Trust Game." *Journal of Economic Behavior & Organization* 43(1):57-73.
Burns, Calvin, Kathryn Mearns, and Peter McGeorge. 2006. "Explicit and Implicit Trust Within Safety Culture." *Risk Analysis* 26(5):1139-1150.
Burt, Ronald S. 1992. *Structural Holes. The Social Structure of Competition*. Cambridge, MA: Harvard University Press.
—. 2003. *Trust, Reputation, and Competitive Advantage*. Oxford: Oxford University Press.
Burt, Ronald S., and Marc Knez. 1995. "Kinds of Third-Party Effects on Trust." *Rationality and Society* 7(3):255-292.
Buskens, Vincent. 1998. "The Social Structure of Trust." *Social Networks* 20(3):265-289.
—. 2002. *Trust and Social Networks*. Boston: Kluwer.
—. 2003. "Trust in Triads: Effects of Exit, Control, and Learning." *Games and Economic Behavior* 42(2):235-252.
Buskens, Vincent, and Werner Raub. 2002. "Embedded Trust: Control and Learning." *Advances in Group Processes* 19:167-202.
—. 2008. "Rational Choice Research on Social Dilemmas: Embededness Effects on Trust." Pp. 1-44 in *Handbook of Rational Choice Research*, edited by R. Wittek, T.A.B. Snijders, and V. Nee. New York: Russell Sage Foundation.
Buskens, Vincent, and Jeroen Weesie. 2000a. "Cooperation via Social Networks." *Analyse & Kritik* 22:44-74.
—. 2000b. "An Experiment on the Effects of Embededdness in Trust Situations: Buying a Used Car." *Rationality and Society* 12(2):228-253.
Butler, J.K., and R.S. Cantrell. 1984. "A Behavioral Decision Theory Approach to Modeling Dyadic Trust in Superiors and Subordinates." *Psychological Reports* 55(1):19-28.
Cacioppo, J.T., and R. E. Petty. 1982. "The Need for Cognition." *Journal of Personality and Social Psychology* 42(1):116-131.
Cacioppo, J.T., R. E. Petty, J. A. Feinstein, and W. B. G. Jarvis. 1996. "Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition." *Psychological Bulletin* 119(2):197-253.
Cacioppo, John T., and Gary G. Berntson. 1994. "Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of Positive and Negative Substrates." *Psychological Bulletin* 115(3):401-423.
Camerer, Colin. 1988. "Gifts as Economic Signals and Social Symbols." *American Journal of Sociology* 94(Supplement):180-214.
—. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
Camerer, Colin F., and Robin M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19(1):7-42.
Camerer, Colin, George Loewenstein, and Drazen Prelec. 2005. "Neuroeconomics: How Neuroscience Can Inform Economics." *Journal of Economic Literature* 43(1):9-64.
Camerer, Colin, and Martin Weber. 1992. "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity." *Journal of Risk and Uncertainty* 5(4):325-370.
Camerer, Colin, and Keith Weigelt. 1988. "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica* 56(1):1-36.
Cameron, Lisa A. 1999. "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia." *Economic Inquiry* 37(1):47-59.
Capra, C. Monica, Kelli Lanier, and Shireen Meer. 2008. "Attitudinal and Behavioral Measures of Trust: A New Comparison." in *Working Paper Series*. Department of Economics, Emory University.
Cassidy, J., and P. R. Shaver. 1999. *Handbook of Attachment: Theory, Research, and Clinical Applications*. New York: Guilford Press.
Chaiken, S., A. Liberman, and A. H. Eagly. 1989. "Heuristic and Systematic Information Processing Within and Beyond the Persuasion Context." Pp. 212-252 in *Unintended Thought: The Limits of Awareness, Intention and Control*, edited by J. S. Uleman, and J. A. Bargh. New York: Guilford Press.

Chaiken, Shelly. 1980. "Heuristic versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion." *Journal of Personality and Social Psychology* 39(5):752-766.
Chaiken, Shelly, and Charles Stangor. 1987. "Attitudes and Attitude Change." *Annual Review of Psychology* 38:575-630.
Chaiken, Shelly, and Yaacov Trope. 1999. *Dual-Process Theories in Social Psychology*. New York, London: The Guilford Press.
Charness, Gary, Ramon Cobo-Reyes, and Natalia Jiménez. 2008. "An Investment Game with Third Party Intervention." *Journal of Economic Behavior & Organization* 68(1):18-28.
Charness, Gary, and Martin Dufwenberg. 2006. "Promises and Partnership." *Econometrica* 74(6):1579-1601.
—. 2007. "Broken Promises." in *Working Paper Series No. 07-17*. Department of Economics, University of Arizona.
Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117(3):817-869.
Chaudhuri, A., and L. Gangadharan. 2002. "Gender Differences in Trust and Reciprocity." in *Working Paper Series, No. 875*. Department of Economics, University of Melbourne.
Chen, Chao C., and Shelly Chaiken. 1999. "The Heuristic-Systematic Model in its Broader Context." Pp. 73-96 in *Dual-Process Theories in Social Psychology*, edited by Shelly Chaiken, and Yaacov Trope. New York, London: Guilford Press.
Chen, Serena, Helen C. Boucher, and Molly P. Tapias. 2006. "The Relational Self Revealed: Integrative Conceptualization and Implications for Interpersonal Life." *Psychological Bulletin* 132(2):151-179.
Chiles, Todd H., and John F: McMackin. 1996. "Integrating Variable Risk Preferences, Trust, and Transaction Cost Economics." *Academy of Management Review* 21(1):73-99.
Clark, M. C., and J. Mills. 1979. "Interpersonal Attraction in Exchange and Communcal Relationships." *Journal of Personality and Social Psychology* 37(1):12-24.
—. 1993. "The Difference Between Communcal and Exchange Relationships." *Personality and Social Psychology Bulletin* 19(6):684-691.
Cleves, Mario A., William W. Gould, and Roberto G. Gutierrez. 2004. *An Introduction to Survival Analysis Using STATA*. College Station, TX: STATA Corporation.
Clore, Gerald L. 1992. "Cognitive Phenomenology: Feelings and the Construction of Judgment." Pp. 133-163 in *The Construction of Social Judgments*, edited by L. L. Martin, and A. Tesser. Hillsdale, NJ: Erlbaum.
—. 2000. "Cognitive Phenomenology: Feelings and the Construction of Judgment." Pp. 133-164 in *The Construction of Social Judgments*, edited by A. Tesser, and L. L. Martin. Hillsdale, NJ: Erlbaum.
Cohen, J., and P. Cohen. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
Cohen, Jacob. 1978. "Partialed Products are Interactions; Partialed Powers are Curve Components." *Psychological Bulletin* 85(4):858-866.
Coleman, James S. 1986. *Interests and Collective Action: Studies in Rationality and Social Change*. Cambridge: Cambridge University Press.
—. 1988. "Social Capital in the Creation of Human Capital." *The American Journal of Sociology* 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure:95-120.
—. 1990. *Foundations of Social Theory*. Cambridge, MA; London, England: Harvard University Press.
Coleman, James S., and Thomas J. Fararo. 1992. *Rational Choice Theory: Adocacy and Critique*. Newbury Park: Sage.
Collins, Nancy L., and Stephen J. Read. 1990. "Adult Attachment, Working Models, and Relationship Quality in Dating Couples." *Journal of Personality and Social Psychology* 58(4):644-663.
Collins, Randall. 1982. *Sociological Insight*. New York: Oxford University Press.
Colombo, Ferdinando, and Guido Merzoni. 2006. "In Praise of Rigidity: The Bright Side of Long-Term Contracts in Repeated Trust Games." *Journal of Economic Behavior & Organization* 59(3):349-373.
Colquitt, Jason A., Brent A. Scott, and Jeffrey A. LePine. 2007. "Trust, Trustworthiness, and Trust Propensity: A Meta-Analytic Test of Their Unique Relationships With Risk Taking and Job Performance." *Journal of Applied Psychology* 92(4):909-927.
Conner, Mark T., Marco Perugini, Rick O´Gorman, Karen Ayres, and Andrew Prestwich. 2007. "Relations Between Implicit and Explicit Measures of Attitudes and Measures of Behavior: Evidence of Moderation by Individual Difference Variables." *Personality and Social Psychology Bulletin* 33(12):1727-1740.
Cook, Karen S. 2001. *Trust in Society*. New York: Russell Sage Foundation.
Cook, Karen S., and Margaret Levi. 1990. *The Limits of Rationality*. Chicago: University of Chicago Press.
Cook, Karen S., Toshio Yamagishi, Coye Cheshire, Robin Cooper, Masafumi Mtsuda, and Rie Mashima. 2005. "Trust Building via Risk Taking: A Cross-Societal Experiment." *Social Psychology Quarterly* 68(2):121-142.

Cookson, R. . 2000. "Framing Effects in Public Good Games." *Experimental Economics* 3(1):55-79.
Cooley, Charles H. 1902. *Human Nature and the Social Order*. New York: Scribner´s.
—. 1983. *Social Organization. A Study of the Larger Mind*. New Brunswick: Transaction Publishers.
Cosmides, George J., and John Tooby. 1992. *Cognitive Adaptations for Social Exchange*. New York: Oxford University Press.
Couch, Laurie L. , Jeffrey M. Adams, and Warren H. Jones. 1996. "The Assessment of Trust Orientation." *Journal of Personality Assessment* 67(2):305-323.
Couch, Laurie L., and Warren H. Jones. 1997. "Measuring Levels of Trust." *Journal of Research in Personality* 31(3):319-336.
Cox, James C. 2002. "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females." Pp. 331-350 in *Experimental Business Research*, edited by Rami Zwick, and Amnon Rapoport. Boston, MA: Springer.
—. 2004. "How to Identify Trust and Reciprocity." *Games and Economic Behavior* 46(2):260-281.
Cronbach, L. J. 1987. "Statistical Tests for Moderator Variables: Flaws in Analyses Recently Proposed." *Psychological Bulletin* 102(3):414-417.
Croson, Rachel T. A., and Nancy R. Buchan. 1999. "Gender and Culture: International Experimental Evidence from Trust Game." *American Economic Review* 89(2):386-391.
Cummings, Larry L., and Philip Bromiley. 1996. "The Organizational Trust Inventory (OTI)." Pp. 302-329 in *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer, and Tom R. Tyler. Thousand Oaks, CA: Sage.
D´Agostino, Ralph B., Albert Belanger, and Ralph B. Jr. D´Agostino. 1990. "A Suggestion for Using Powerful and Informative Tests of Normality." *The American Statistician* 44(4):316-321.
D´Andrade, Roy G. 1995. *The Development of Cognitive Anthropology*. New York: Cambridge University Press.
Daft, Richard L., and Robert H. Lengel. 1984. "Information Richness: A New Approach to Manager Information Processing and Organization Design." Pp. 191-233 in *Research in Organizational Behavior*, edited by Barry M. Staw, and Larry L. Cummings. Greenwhich: JAI Press.
—. 1986. "Organizational Information Requirements, Media Richness, and Structural Design." *Management Science* 32(5):554-571.
Damasio, Antonio R. 1994. *Descartes' Error: Emotion, Reason and the Human Brain*. London: Picador.
Das, T. K., and Bing-Sheng Teng. 1998. "Between Trust and Control: Developing Confidence in Partner Cooperation in Alliances." *Academy of Management Review* 23(3):491-512.
—. 2001. "Trust, Control, and Risk in Strategic Alliances: An Integrated Framework." *Organization Studies* 22(2):251-283.
Dasgupta, Partha. 1988. "Trust as a Commodity." Pp. 49-72 in *Trust: Making and Breaking Cooperative Relations*, edited by Diego Gambetta. Oxford, New York: Blackwell Publishing.
Dawes, R., J. McTavish, and H. Shaklee. 1977. "Behavior, Communicaiton, and Assumptions About Other People´s Behavior in a Commons Dilemma Situation." *Journal of Personality and Social Psychology* 35(1):1-11.
De Cremer, David, and Mark van Vugt. 1999. "Social Identification Effects in Social Dilemmas: A Transformation of Motives." *European Journal of Social Psychology* 29:871-893.
DeNeve, Kristina M., and Harris Cooper. 1998. "The Happy Personality: A Meta-Analysis of 137 Personality Traits and Subjective Well-Being." *Psychological Bulletin* 124(2):197-229.
DePaulo, Bella M. 1992. "Nonverbal Behavior and Self-Presentation." *Psychological Bulletin* 11(2):203-243.
Deutsch, Morton. 1958. "Trust and Suspicion." *The Journal of Conflict Resolution* 2(4):265-279.
—. 1960. "Trust, Trustworthiness and the F-Scale." *Journal of Abnormal and Social Psychology* 61(1):138-140.
—. 1973. *The Resolution of Conflict. Constructive and Destructive Processes*. New Haven, London: Yale University Press.
Devine, P. G. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled Components." *Journal of Personality and Social Psychology* 56(1):5-18.
Devine, P. G., E. A. Plant, D. M. Amodio, E. Harmon-Jones, and S. L. Vance. 2002. "The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Responds Without Prejudice." *Journal of Personality and Social Psychology* 82(5):835-848.
Diekmann, Andreas. 2004. "The Power of Reciprocity: Fairness, Reciprocity, and Stakes in Variants of the Dictator Game." *Journal of Conflict Resolution* 48(4):487-505.
Diekmann, Andreas, and Peter Preisendörfer. 1992. "Persönliches Umweltverhalten. Diskrepanzen zwischen Anspruch und Wirklichkeit." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 44:226-251.
—. 2003. "Green and Greenback: The Behavioral Effects of Environmental Attitudes in Low-Cost and High-Cost Situations." *Rationality and Society* 15(4):441-472.
Dietz, Graham, Nicole Gillespie, and Georgia T. Chao. 2010. "Unravelling the Complexities of Trust and Culture." Pp. 3-41 in *Organizational Trust. A Cultural Perspective*, edited by Mark N. Saunders, Denise

Skinner, Graham Dietz, Nicole Gillespie, and Roy J. Lewicki. Cambridge (a.o.): Cambridge University Press.
Dijksterhuis, A., and J. A. Bargh. 2001. "The Perception-Behavior Expressway: Automatic Effects of Social Perception on Social Behavior." Pp. 1-40 in *Advances in Experimental Social Psychology, Vol 33*. San Diego: Academic Press Inc.
Dillard, J. P., D.H. Solomon, and M. T. Palmer. 1999. "Structuring the Concept of Relational Communication." *Communication Monographs* 66(1):49-65.
Dillard, J. P., D.H. Solomon, and J. A. Samp. 1996. "Framing Social Reality: The Relevance of Relational Judgments." *Communication Research* 23(6):703-723.
DiMaggio, Paul, and Walter. W. Powell. 1991. "Introduction." Pp. 1-38 in *The New Institutionalism in Organizational Analysis*, edited by Paul DiMaggio, and Walter. W. Powell. Chicago: University of Chicago Press.
Dirks, Kurt T., and Donald L. Ferrin. 2001. "The Role of Trust in Organizational Settings." *Organization Science* 12(4):450-467.
Dirks, Kurt T., Roy J. Lewicki, and Akbar Zaheer. 2009. "Repairing Relationships Within and Between Organizations: Building a Conceptual Foundation." *Academy of Management Review* 34(1):68-84.
Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2008. "The Intergenerational Transmission of Risk and Trust Attitudes." in *CESifo Working Paper Series, No. 2307*. Munich Society for the Promotion of Economic Research (CESifo), University of Munich.
Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3):522-550.
Dolan, Conor V., Han L. J. Van der Maas, and Peter C. M. Molenaar. 2002. "A Framework for ML Estimation of Parameters of (Mixtures of) Common Reaction Time Distributions Given Optimal Truncation or Censoring." *Behavior Research Methods* 34(3):304-323.
Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt. 2011. "The Framing of Games and the Psychology of Play." *Games and Economic Behavior* 73(2):459-478.
Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior* 47(2):268-289.
Dunn, Jennifer R., and Maurice E. Schweitzer. 2005. "Feeling and Believing: The Influence of Emotion on Trust." *Journal of Personality and Social Psychology* 88(5):736-748.
Durkheim, Emile. 1984. *The Division of Labour in Society*. Basingstoke: Macmillan.
Eagly, A. H., R. D. Ashmore, M. G. Makhijani, and L. C. Longo. 1991. "What is Beautiful is Good, But: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype." *Psychological Bulletin* 110(1):109-128.
Eagly, A. H., and S. Chaiken. 1993. "The Psychology of Attitudes." Fort Worth: Harcourt Brace Jovanovich.
Echambadi, Raj, Inigo Arroniz, Werner Reinartz, and Junsoo Lee. 2006. "Empirical Generalizations From Brand Extension Research: How Sure Are We?" *International Journal of Research in Marketing* 23(3):253-261.
Echambadi, Raj, and James D. Hess. 2007. "Mean-Centering Does Not Alleviate Multicollinearity Problems in Moderated Multiple Regression Models." *Marketing Science* 26(3):438-445.
Eckel, Catherine C., and Philip J. Grossman. 2005. "Managing Diversity by Creating Team Identity." *Journal of Economic Behavior & Organization* 58(3):371-392.
Eckel, Catherine C., and Rick K. Wilson. 2003. "The Human Face of Game Theory: Trust and Reciprocity in Sequential Games." Pp. 245-274 in *Trust and Reciprocity. Interdisciplinary Lessons From Experimental Research.*, edited by Elinor Ostrom, and James Walker. New York: Russel Sage Foundation.
—. 2004. "Is Trust a Risky Decision?" *Journal of Economic Behavior & Organization* 55(4):447-465.
Edwards, Ward. 1954. "The Theory of Decision Making." *Psychological Bulletin* 51(4):380-417.
Einhorn, Hillel J., and Robin M. Hogarth. 1986. "Decision Making Under Ambiguity." *The Journal of Business* 59(4):225-250.
Ekman, P. 1972. "Universals and Cultural Differences in Facial Expressions of Emotions." in *Nebraska Symposium on Motivation*, edited by J. K. Cole. Lincoln: University of Nebraska Press.
Ellingsen, Tore, Magnus Johanneson, Sigve Tjotta, and Gaute Torsvik. 2010. "Testing Guilt Aversion." *Games and Economic Behavior* 68(1):95-107.
Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *The Quarterly Journal of Economics* 75(4):643-669.
Elster, Jon. 1979. *Ulysses and the Sirens. Studies in Rationality and Irrationality*. Cambridge, New York: Cambridge University Press.
—. 1982. "Marxism, Functionalism, and Game Theory." *Theory and Society* 11(4):453-482.

—. 1986a. "Introduction." Pp. 1-33 in *Rational Choice. Readings in Social and Political Theory*, edited by Jon Elster. Oxford: Blackwell.
—. 1986b. *Rational Choice. Readings in Social and Political Theory*. New York: New York University Press.
—. 1989. *The Cement of Society. A Study of Social Order*. Cambridge: Cambridge University Press.
—. 2005. "Fehr on Altruism, Emotion, and Norms." *Analyse & Kritik* 27(1):197-211.
Emirbayer, Mustafa, and Ann Mische. 1998. "What is Agency?" *American Journal of Sociology* 103(4):962-1023.
Endress, Martin. 2001. "Vertrauen und Vertrautheit - Phänomenologisch-anthropologische Grundlegung." Pp. 161-203 in *Vertrauen - Die Grundlage des sozialen Zusammenhalts*, edited by Martin Hartmann, and Claus Offe. Frankfurt a. Main, New York: Campus.
—. 2002. *Vertrauen*. Bielefeld: transcript.
Engle-Warnick, Jim, and Robert L. Slonim. 2004. "The Evolution of Strategies in the Repeated Trust Game." *Journal of Economic Behavior & Organization* 55(4):553-573.
—. 2006. "Learning to Trust in Indefinitely Repeated Games." *Games and Economic Behavior* 54(1):95-114.
Ensminger, Jean. 2001. "Reputations, Trust, and the Principal Agent Problem." Pp. 158-201 in *Trust in Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Epstein, Seymour. 1991. "Cognitive-Experiential Self-Theory: An Integrative Theory of Personality." Pp. 111-137 in *The Self with Others: Convergences in Psychoanalytical, Social and Personality Psychology*, edited by R. Curtis. New York: Guilford Press.
Epstein, Seymour, Rosemary Pacini, Veronika Denes-Raj, and Harriet Heier. 1996. "Individual Differences in Intuitive-Experiential and Analytical-Rational Thinking Styles." *Journal of Personality and Social Psychology* 71(2):390-405.
Erber, M. W., S. D. Hodges, and T. D. Wilson. 1995. "Attitude Strength, Attitude Stability, and the Effects of Analyzing Reasons." Pp. 433-454 in *Attitude Strength: Antecedents and Consequences*, edited by R. E. Petty, and J. A. Krosnick. Mahwah, NJ: Lawrence Erlbaum Associates.
Erikson, Erik H. 1950. *Childhood and Society*. New York: Norton.
—. 1968. *Identity: Youth and Crisis*. New York: Norton.
—. 1989. *Identität und Lebenszyklus*, 11 ed. Frankfurt am Main: Suhrkamp.
Ermisch, John, and Diego Gambetta. 2010. "Do Strong Family Ties Inhibit Trust?" *Journal of Economic Behavior & Organization* 75(3):365-376.
Ermisch, John, Diego Gambetta, Heather Laurie, Thomas Siedler, and Noah S. C. Uhrig. 2009. "Measuring People´s Trust." *Journal of the Royal Statistical Society, Series A* 172(4):749-769.
Esser, Hartmut. 1990. "'Habits', 'Frames' und 'Rational Choice'." *Zeitschrift für Soziologie* 19(4):231-247.
—. 1991. "Die Rationalität des Alltagshandelns." *Zeitschrift für Soziologie* 20(6):430-445.
—. 1993a. "Social Modernization and the Increase in the Divorce Rate." *Journal of Institutional and Theoretical Economics* 149(1):252-277.
—. 1993b. *Soziologie. Allgemeine Grundlagen*, 1 ed. Frankfurt a. Main (a.o.): Campus.
—. 1999a. *Soziologie. Allgemeine Grundlagen*, 3 ed. Frankfurt a. Main (a.o.): Campus.
—. 1999b. *Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt a. Main (a.o.): Campus.
—. 2000a. *Soziologie. Spezielle Grundlagen. Band 3: Soziales Handeln*. Frankfurt a. Main (a.o.): Campus.
—. 2000b. *Soziologie. Spezielle Grundlagen. Band 4: Opportunitäten und Restriktionen*. Frankfurt a. Main (a.o.): Campus.
—. 2000c. *Soziologie. Spezielle Grundlagen. Band 5: Institutionen*. Frankfurt a. Main (a.o.): Campus.
—. 2001. *Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur*. Frankfurt a. Main (a.o.): Campus.
—. 2002. "In guten wie in schlechten Tagen? Das Framing der Ehe und das Risiko zu Scheidung. Eine Anwendung und ein Test des Modells der Frame-Selektion." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54(1):27-63.
—. 2005. "Affektuelles Handeln: Emotionen und das Modell der Frame-Selektion." in *Working Paper Series, No. 05-15*. Sonderforschungsbereich 504: Rationalitätskonzepte, Entscheiduntgsverhalten und ökonomische Modellierung (SFB 504): Unversity of Mannheim.
—. 2009. "Rationality and Commitment: The Model of Frame-Selection and the Explanation of Normative Action." Pp. 207-230 in *Raymond Boudon: A Life in Sociology, Vol 2, Part Two: Toward a General Theory of Rationality*, edited by Mohamed Cherkauoi, and Peter Hamilton. Oxford: The Bardwell Press.
—. 2010. "Das Modell der Frame-Selektion. Eine allgemeine Handlungstheorie für die Sozialwissenschaften?" *Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft* 50:45-62.
Evans, J. S. B. T. 2008. "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition." *Annual Review of Psychology* 59:255-278.

—. 2009. "How Many Dual-Process Theories Do We Need? One, Two, Many?" Pp. 33-54 in *In Two Minds. Dual Processes and Beyond*, edited by J. S. B. T. Evans, and K. Frankish. New York: Oxford University Press.
Evans, J. S. B. T., and K. Frankish. 2009. *In Two Minds. Dual Processes and Beyond*. New York: Oxford University Press.
Fairclough, Stephen, and Ben Mulder. 2011. "Psychophysiological Processes of Mental Effort Investment." in *Motivation Perspectives on Cardiovascular Response: Mechanisms and Applications*, edited by R. Wright, and G. Gendolla (in press).
Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing Theories of Fairness - Intentions Matter." *Games and Economic Behavior* 62(1):287-303.
Falk, Armin, and Urs Fischbacher. 2000. "A Theory of Reciprocity." in *Working Paper Series, No. 6*. Institute for Empirical Research in Economics, University of Zurich.
—. 2006. "A Theory of Reciprocity." *Games and Economic Behavior* 54:293-315.
Falk, Armin, and James J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326:535-538.
Farrell, Henry. 2004. "Trust, Distrust, and Power." Pp. 85-105 in *Distrust*, edited by Russell Hardin. New York: Russel Sage Foundations.
Farrell, Joseph. 1987. "Cheap Talk, Coordination, and Entry." *The RAND Journal of Economics* 18(1):34-39.
Fazio, R.H. 1986. "How do Attitudes Guide Behavior?" Pp. 204-243 in *Handbook of Motivation and Cognition*, edited by R.M. Sorrentino, and E. T. Higgins. New York: Guilford Press.
—. 1990a. "Multiple Processes by which Attitudes Guide Behavior: the MODE Model as an Integrative Framework." *Advances in Experimental Social Psychology* 23:75-109.
—. 1990b. "A Practical Guide to the Use of Response Latency in Social Psychological Research." Pp. 74-97 in *Rsearch Methods in Personality and Social Psychology*, edited by Hendrick Clyde, and Margaret S. Clark. Thousan Oaks, CA: Sage.
—. 1995. "Attitudes as Object-Evaluation Associations: Determinants, Consequences, and Correlates of Attitude Accessibility." Pp. 247-282 in *Attitude Strength: Antecedents and Consequences*, edited by R. E. Petty, and J. A. Krosnick. Hillsdale, NJ: Lawrence Erlbaum.
—. 2001. "On the Automatic Effect of Associated Evaluations: An Overview." *Cognition and Emotion* 15(2):115-141.
—. 2007. "Attitudes as Object-Evaluation Associations of Varying Strength." *Social Cognition* 25(5):603-637.
Fazio, R.H., and Tamara Towles-Schwen. 1999. "The MODE Model of Attitude-Behavior Processes." Pp. 97-116 in *Dual-Process Theories in Social Psychology*, edited by Shelly Chaiken, and Yaacov Trope. New York, London: The Guilford Press.
Fazio, R.H., and C. J. Williams. 1986. "Attitude Accessibility as a Moderator of the Attitude-Perception and Attitude-Behavior Relations: An Investigation of the 1984 Presidential Election." *Journal of Personality and Social Psychology* 51(3):505-514.
Fehr, Ernst. 2008. "On the Economics and Biology of Trust." in *IZA Discussion Paper Series, No. 3895*. Bonn: Institute for the Study of Labor (IZA).
—. 2009. "On the Economics and Biology of Trust." *Journal of the European Economic Association* 7(2):235-266.
Fehr, Ernst, Martin Brown, and Christian Zehnder. 2008. "On Reputation: A Microfoundation of Contract Enforcement and Price Rigidity." in *Swiss National Bank Working Papers, No. 2008-17*. Zurich: Swiss National Bank.
Fehr, Ernst, and Colin Camerer. 2007. "Social Neuroeconomics: The Neural Circuitry of Social Preferences." *Trends in Cognitive Sciences* 11(10):419-427.
Fehr, Ernst, and Urs Fischbacher. 2002. "Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives." *The Economic Journal* 112(3):1-33.
—. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior* 25:63-87.
Fehr, Ernst, Urs Fischbacher, and Michael Kosfeld. 2005. "Neuroeconomic Foundations of Trust and Social Preferences: Initial Evidence." *American Economic Review* 95(2):346-351.
Fehr, Ernst, Urs Fischbacher, and Elena Tougareva. 2002a. "Do High Stakes and Competition Undermine Fairness? Evidence from Russia." in *Working Paper Series, No. 120*. Institute for Empirical Research in Economics, University of Zurich.
Fehr, Ernst, Urs Fischbacher, Bernhard von Rosenbladt, Jürgen Schupp, and Gert G. Wagner. 2002b. "A Nationwide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys." in *Working Paper Series, No. 141*. Institute for Empirical Research in Economics, University of Zurich.
Fehr, Ernst, and Simon Gächter. 2000a. "Cooperation and Punishment in Public Goods Games." *American Economic Review* 90(4):980-994.

—. 2000b. "Fairness and Retaliation: The Economics of Reciprocity." *The Journal of Economic Perspectives* 14(3):159-181.
—. 2002a. "Altruistic Punishment in Humans." *Nature* 415(6868):137-140.
—. 2002b. "Do Incentive Contracts Undermine Voluntary Cooperation?" in *Working Paper Series, No. 34*. Institute for Empirical Research in Economics, University of Zürich.
Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. "Reciprocity as a Contractural Enforcement Device: Experimental Evidence." *Econometrica* 65(4):833-860.
Fehr, Ernst, and Herbert Gintis. 2007. "Human Motivation and Social Cooperation: Experimental and Analytical Foundations." *Annual Review of Sociology* 33:43-64.
Fehr, Ernst, and Karla Hoff. 2011. "Tastes, Castes and Culture: The Influence of Society in Preferences." in *IZA Discussion Paper Series, No. 5919*. Institute for the Study of Labor (IZA), Bonn.
Fehr, Ernst, and John A. List. 2004. "The Hidden Costs and Returns of Incentives - Trust and Trustworthiness among CEOs." *Journal of the European Economic Association* 2(5):743-771.
Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114(3):817-868.
—. 2006. "The Economics of Fairness, Reciprocity and Altruism, Experimental Evidence and New Theories." Pp. 615-684 in *Handbook of the Economics of Giving, Altruism and Reciprocity, Vol. 1*, edited by Serge-Christoph Kolm, and Jean Mercier Ythier. Amsterdam: Elsevier.
Fehr, Ernst, and Jean-Robert Tyran. 2008. "Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia." *Econometrica* 76(2):353-394.
Feingold, A. 1992. "Good-Looking People are not What We Think." *Psychological Bulletin* 111(2):304-341.
Ferrin, Donald L., Michelle C. Bligh, and Jeffrey C. Kohles. 2008. "It takes Two to Tango: An Interdependence Analysis of the Spiraling of Perceived Trustworthiness and Cooperation in Interpersonal and Intergroup Relationships." *Organizational Behavior and Human Decision Processes* 107(2):161-178.
Fischbacher, U. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10(2):171-178.
Fiske, S. T. 1982. "Schema-Triggered Affect: Applications to Social Perception." Pp. 55-78 in *Affect and Cognition: The 17th Annual Carnegie Symposium on Cognition*, edited by M. C. Clark, and S. T. Fiske. Hillsdale, NJ: Erlbaum.
—. 1991. *Structures of Social Life*. New York: Free Press.
—. 2004. *Social Beings: A Core Motives Approach to Social Psychology*. New York: John Wiley & Sons.
Fiske, S. T., M. Lin, and S. L. Neuberg. 1999. "The Continuum Model. Ten Years Later." Pp. 231-254 in *Dual-Process Theories in Social Psychology*, edited by Shelly Chaiken, and Yaacov Trope. New York, London: Guilford Press.
Fiske, S. T., and S. L. Neuberg. 1990. "A Continuum of Impression Formation, from Category-based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation." Pp. 1-74 in *Advances in Experimental Social Psychology*, edited by M.P. Zanna. New York: Academic Press.
Fiske, S. T., and M. Pavelchak. 1986. "Category-based Versus Piecemeal-based Affective Responses: Developments in Schema-Triggered Affect." Pp. 167-203 in *Handbook of Motivation and Cognition*, edited by R.M. Sorrentino, and E. T. Higgins. New York: Guilford Press.
Fiske, Susan T. 1993. "Social Cognition and Social Perception." *Annual Review of Psychology* 44:155-194.
Fiske, Susan T., and Shelley E. Taylor. 1991. *Social Cognition*. New York: McGraw-Hill.
Florack, Arnd, Martin Scarabis, and Herbert Bless. 2001. "When Do Associations Matter? The Use of Automatic Associations Toward Ethnic Groups in Person Judgments." *Journal of Experimental Social Psychology* 37(6):518-524.
Forgas, Joseph P. 2002. "Feeling and Doing: Affective Influences on Interpersonal Behavior." *Psychological Inquiry* 13(1):1-28.
Forgas, Joseph P., and Rebekah East. 2008. "On Being Happy and Gullible: Mood Effects on Skepticism and the Detection of Deception." *Journal of Experimental Social Psychology* 44(5):1362-1367.
Förster, Jens, and Markus Denzler. 2009. "A Social-Cognitive Perspective on Automatic Self-Regulation. The Relevance of Goals in the Information-Processing Sequence." Pp. 245-268 in *Social Cognition. Basis of Human Interaction*, edited by Fritz Strack, and Jens Förster. New York, London: Psychology Press.
Förster, Jens, Nira Liberman, and Ronald S. Friedman. 2007. "Seven Principles of Goal Activation: A Systematic Approach to Distinguishing Goal Priming From Priming of Non-Goal Constructs." *Personality and Social Psychology Review* 11(3):211-233.
Fraley, Chris R. 2002. "Attachment Stability from Infancy to Adulthood: Meta-Analysis and Dynamic Modeling of Developmental Mechanisms." *Personality and Social Psychology Review* 6(1):123-151.
—. 2010. "A Brief Overview of Adult Attachment Theory and Research." University of Illinois, Department of Psychology. URL: http://internal.psychology.illinois.edu/~rcfraley/attachment.htm (Last Accessed 05/06/2011).

Fraley, Chris R., and Susan J. Spieker. 2003. "Are Infant Attachment Patterns Continously or Categorically Distributed? A Taxometric Analysis of Strange Situation Behavior." *Developmental Psychology* 39(3):387-404.
Frederick, Shane, George Loewenstein, and Ted O´Donoghue. 2002. "Time Discounting and Time Preferences: A Critical Review." *Journal of Economic Literature* 40(2):351-401.
Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15(5):589-611.
Friedman, Debra, and Michael Hechter. 1988. "The Contribution of Rational Choice Theory to Macrosociological Research." *Sociological Theory* 6(2):201-218.
Frijda, Nico H. 1988. "The Laws of Emotion." *American Psychologist* 43(5):349-358.
Frisch, Deborah, and Jonathan Baron. 1988. "Ambiguity and Rationality." *Journal of Behavioral Decision Making* 1(3):149-157.
Fudenberg, Drew, and Eric Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54(3):533-554.
Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge, MA: MIT Press.
Fukuyama, Francis. 1995. *Trust. The Social Virtues and the Creation of Prosperity*. New York: The Free Press.
Gabarro, J.J. 1978. "The Development of Trust, Influence and Expectations." Pp. 290-203 in *Interpersonal Behavior: Communication and Understanding in Relationships*, edited by A.G. Athos, and J.J. Gabarro. Englewood Cliffs, NJ: Prentice Hall.
Gächter, Simon, Benedikt Herrmann, and Christian Thöni. 2004. "Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence." *Journal of Economic Behavior & Organization* 55(4):505-531.
Gambetta, Diego. 1988a. "Can We Trust Trust?" Pp. 213-237 in *Trust. Making and Braking Cooperative Relations.*, edited by Diego Gambetta. New York: Basil Blackwell.
—. 1988b. *Trust: Making and Braking of Cooperative Relations*. New York: Basil Blackwell.
—. 1993. *The Sicilian Mafia: The Business of Private Protection*. Cambridge, MA; London, England: Harvard University Press.
Ganesan, Shankar. 1994. "Determinants of Long-Term Orientation in Buyer-Seller Relationships." *The Journal of Marketing* 58(2):1-19.
Garbarino, Ellen, and Robert Slonim. 2009. "The Robustness of Trust and Reciprocity Across a Heterogeneous U.S Population." *Journal of Economic Behavior & Organization* 69(3):226-240.
Garfinkel, Harold. 1963. "A Conception of and Experiments with 'Trust' as a Condition of Stable Concerted Actions." in *Motivation and Social Interaction: Cognitive Determinants*, edited by O.J. Harvey. New York: Ronald Press.
—. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, New Jersey: Prentice Hall.
Geneakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological Games and Sequential Rationality." *Games and Economic Behavior* 1(1):60-80.
Gibbons, Robert. 2001. "Trust in Social Structures: Hobbes and Coase Meet Repeated Games." Pp. 332-353 in *Trust in Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Giddens, Anthony. 1984. *The Constitution of Society. Outline of the Theory of Structuration*. Berkeley: University of California Press.
—. 1990. *The Consequences of Modernity*. Stanford, CA: Stanford University Press.
—. 1991. *Modernity and Self-Identity*. Cambridge: Polity Press.
—. 1994. "Risk, Trust, Reflexivity." Pp. 184-197 in *Reflexive Modenization*, edited by Ulrich Beck, Anthony Giddens, and Scott Lash. Cambridge: Polity Press.
Gigerenzer, Gerd. 2000. *Adaptive Thinking: Rationality in the Real World*. Oxford: Oxford University Press.
Gigerenzer, Gerd, and Wolfgang Gaissmaier. 2011. "Heuristic Decision Making." *Annual Review of Psychology* 62:451-482.
Gigerenzer, Gerd, and Daniel G. Goldstein. 1996. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* 103(4):650-669.
Gigerenzer, Gerd, and Reinhard Selten. 2001. *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.
Gill, Harjinder, Kathleen Boies, Joan E. Finegan, and Jeffrey McNally. 2005. "Antecedents of Trust: Establishing a Boundary Condition for the Relation Between Propensity to Trust and Intention to Trust." *Journal of Business and Psychology* 19(3):287-302.
Gintis, Herbert. 2000a. *Game Theory Evolving*. Princeton, NJ: Princeton University Press.
—. 2000b. *Game Theory Evolving. A Problem-Centered Introduction to Modeling*
Strategic Interaction. Princeton: Princeton University Press.
—. 2000c. "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology* 206(2):169-179.

—. 2007. "A Framework for the Unification of the Behavioral Sciences." *Behavioral and Brain Sciences* 30(1):1-61.
Gintis, Herbert, Samuel Bowles, and Ernst Fehr. 2003. "Explaining Altruisitc Behavior in Humans." *Evolution and Human Behavior* 24(3):153-229.
Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter. 2000. "Measuring Trust." *Quarterly Journal of Economics* 115(3):811-846.
Glöckner, Andreas, and Cilia Witteman. 2010. "Beyond Dual-Process Models: A Categorisation of Processes Underlying Intuitive Judgment and Decision Making." *Thinking & Reasoning* 16(1):1-25.
Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Harmondsworth, UK: Penguin.
—. 1967. *Interaction Ritual: Essays in Face-to-Face Behavior*. Chicago: Aldine.
—. 1974. *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.
Goldberg, Jeffrey, Livia Markoczy, and Lawrence G. Zahn. 2005. "Symmetry and the Illusion of Control as Bases for Cooperative Behavior." *Rationality and Society* 17(2):243-270.
Good, David. 1988. "Individuals, Interpersonal Relations, Trust." Pp. 31-48 in *Trust. Making and Braking Cooperative Relations*, edited by Diego Gambetta. New York: Blackwell.
Gouldner, Alvin W. 1960. "The Norm of Reciprocity: A Preliminary Statement." *American Sociological Review* 25(2):161-178.
Govier, Trudy. 1993. "Self-Trust, Autonomy, and Self-Esteem." *Hypatia* 8(1):99-120.
Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *The American Journal of Sociology* 91(3):481-510.
Green, Donald P., and Ian Shapiro. 1994. *The Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven: Yale University Press.
Greenwald, Anthony G., and Mazharin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102(1):4-27.
Greenwald, Anthony G., and Anthony R. Pratkanis. 1984. "The Self." Pp. 129-178 in *Handbook of Social Cognition*, edited by R. S. Wyer, and T. K. Srull. Hillsdale, NJ: Erlbaum.
Greifeneder, Rainer, Herbert Bless, and Michel Tuan Pham. 2011. "When Do People Rely on Affective and Cognitive Feelings in Judgment? A Review." *Personality and Social Psychology Review* 15(2):107-141.
Greifeneder, Rainer, Patrick Müller, Dagmar Stahlberg, Kees Van Den Bos, and Herbert Bless. 2010. "Guiding Trustful Behavior: The Role of Accessible Content Versus Accessible Experiences." *Journal of Behavioral Decision Making* 24(5):498-514.
Guerra, Gerardo, and Daniel John Zizzo. 2004. "Trust Responsiveness and Beliefs." *Journal of Economic Behavior & Organization* 55(1):25-30.
Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "The Role of Social Capital in Financial Development." *American Economic Review* 94(3):526-556.
—. 2008. "Long Term Persistence." in *NBER Working Paper Series, No. 14278*. National Bureau of Economic Research, Cambridge, MA.
—. 2009. "Cultural Biases in Economic Exchange?" *Quarterly Journal of Economics* 124(3):1095-1131.
Gurtman, Michael B. 1992. "Trust, Distrust, and Interpersonal Problems: A Circumplex Analysis." *Journal of Personality and Social Psychology* 62(6):989-1002.
Güth, W., M. Levati, and M. Ploner. 2008. "Social Identity and Trust. An Experimental Investigation." *Journal of Socio-Economics* 37(4):1293-1308.
Güth, Werner, Carsten Schmidt, and Matthias Sutter. 2005. "Bargaining Outside the Lab: A Newspaper Experiment of a Three-Person Ultimatum Game." *The Economic Journal* 117(518):449-469.
Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108(4):814-834.
Hardin, Russell. 1982. *Collective Action*. Baltimore: Johns Hopkins University Press.
—. 1993. "The Street-Level Epistemology of Trust." *Politics & Society* 21(4):505-529.
—. 2001. "Conceptions and Explanations of Trust." Pp. 3-39 in *Trust in Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
—. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.
—. 2003. "Gaming Trust." Pp. 80-101 in *Trust and Reciprocity. Intersdiciplinary Lessons from Experimental Research*, edited by Elinor Ostrom, and James Walker. New York: Russel Sage Foundation.
Harrison, Glenn W., M.I. Lau, and E.E. Rutstrom. 2007. "Estimating Risk Attitudes in Denmark: A Field Experiment." *Scandinavian Journal of Economics* 109(2):341-368.
Harsanyi, John C. 1967, 1968. "Games With Incomplete Information Played by 'Bayesian' Players I-III." *Management Science* 14(159-182, 320-334, 486-502).
Haselton, Martie G., and Daniel Nettle. 2006. "The Paranoid Optimist: An Integrative Model of Cognitive Biases." *Personality and Social Psychology Review* 10(1):47-66.

Haxby, James V., Elizabeth A. Hoffman, and Ida M. Gobbini. 2002. "Human Neural Systems for Face Recognition and Social Communication." *Biological Psychiatry* 51:59-67.
Hazan, Cindy, and Philip Shaver. 1987. "Romantic Love Conceptualized as an Attachment Process." *Journal of Personality and Social Psychology* 52(3):511-524.
Heathcote, A., S. Brown, and D. Cousineau. 2004. "QMPE: Estimating Lognormal, Wald, and Weibull RT Distributions With a Parameter-Dependent Lower Bound." *Behavior Research Methods* 36(2):277-290.
Heathcote, A., S. J. Popiel, and D. J. Mehwort. 1991. "Analysis of Response Time Distributions: An Example Using the Stroop Task." *Psychological Bulletin* 109(2):340-347.
Hedström, Peter, and Richard Swedberg. 1996. "Rational Choice, Empirical Research, and the Sociological Tradition." *European Sociological Review* 12(2):127-146.
—. 1998. *Social Mechanisms. An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
Heimer, Carol A. 2001. "Solving the Problem of Trust." Pp. 40-87 in *Trust and Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Heiner, R. A. 1985. "Origin of Predictable Behavior - Further Modeling and Applications." *American Economic Review* 75(2):391-396.
Heiner, Ronald A. 1983. "The Origin of Predictable Behavior." *American Economic Review* 73(4):560-595.
Helliwell, John F., and Robert D. Putnam. 2004. "The Social Context of Well-Being." *Philosophical Transactions of the Royal Society London* 359:1435-1446.
Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15(2):135-175.
Henslin, James M. 1968. "Trust and the Cab Driver." Pp. 139-159 in *Sociology and Everyday Life*, edited by Marcello Truzzi. Englewood Cliffs, NJ: Prentice-Hall.
Hertwig, Ralph, and Andreas Ortmann. 2001. "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *Behavioral and Brain Sciences* 24(3):282-451.
Higgins, E. T. 1996. "Knowledge Activation: Accessibility, Applicability, and Salience." Pp. 133-168 in *Social Psychology: Handbook of Basic Principles*, edited by E. T. Higgins, and A. Kruglanski. New York: Guilford Press.
Higgins, E. T., G. A. King, and G. H. Mavin. 1982. "Individual Construct Accessibility and Subjective Impressions and Recall." *Journal of Personality and Social Psychology* 43(1):35-47.
Hill, Claire, and Erin A. O'Hara. 2006. "A Cognitive Theory of Trust." *Washington University Law Review* 84(7):1717-1796.
Hoffman, E., K. McCabe, and V. L. Smith. 1996. "On Expectations and the Monetary Stakes in Ultimatum Games." *International Journal of Game Theory* 25(3):289-301.
Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 2008. "Social Distance and Reciprocity in Dictator Games." Pp. 411-415 in *Handbook of Experimental Economics Results, Vol. 1*, edited by Charles R. Plott, and Vernon L. Smith. Amsterdam: Elsevier.
Hogarth, Robin M., and Melvin W. Reder. 1987. *Rational Choice: The Contrast between Economics and Psychology*. Chicago: University of Chicago Press.
Hogg, M. 2006. "Social Identity Theory." Pp. 111-136 in *Contemporary Social Psychological Theories*, edited by Peter J. Burke. Stanford: Stanford University Press.
Hohle, Raymond H. 1965. "Inferred Components of Reaction Times as Functions of Foreperiod Duration." *Journal of Experimental Psychology* 69(4):382-386.
Hollingshead, A. B. 1996. "Information Suppression and Status Persistence in Group Decision Making: The Effects of Communication Media." *Human Communication Research* 23(3):193-219.
Holm, Hakan, and Paul Nystedt. 2008. "Trust in Surveys and Games – A Methodological Contribution on the Influence of Money and Location." *Journal of Economic Psychology* 29(4):522-542.
Holmes, J. G. 1991. "Trust and the Appraisal Process in Close Relationships." Pp. 57-104 in *Advances in Personal Relationships, Vol. 2*, edited by Warren H. Jones, and Daniel Perlman. Oxford, England: Jessica Kingsley Publishers.
Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5):1644-55.
Hosmer, Larue T. 1995. "The Connecting Link Between Organizational Theory and Philosophical Ethics." *The Academy of Management Review* 20(2):379-403.
Houser, Daniel, Erte Xiao, Kevin McCabe, and Vernon L. Smith. 2008. "When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation." *Games and Economic Behavior* 62(2):509-532.
Huang, Li, and J. Keith Murnighan. 2010. "What´s in a Name? Subliminally Activating Trusting Behavior." *Organizational Behavior and Human Decision Processes* 111(1):62-70.

Hunkler, Christian, and Thorsten Kneip. 2008. "Das Zusammenspiel von Normen und Anreizen bei der Erklärung partnerschaftlicher Stabilität." in *MZES Working Paper Nr. 108/2008*. Mannheimer Zentrum für empirische Sozialforschung (MZES), Universität Mannheim.
Isaac, R. Mark, and James M. Walker. 1988. "Communicatiom and Free-Riding Behavior: The Voluntary Contribution Mechanism." *Economic Inquiry* 26(4):585-608.
Jaccard, James, and Robert Turrisi. 2003. *Interaction Effects in Multiple Regression*, 2 ed. Thousand Oaks, London: Sage Publications.
James, Harvey S. 2002a. "On the Reliability of Trusting." *Rationality and Society* 14(2):229-256.
—. 2002b. "The Trust Paradox: A Survey of Economic Inquiries Into the Nature of Trust and Trustworthiness." *Journal of Economic Behavior & Organization* 47(3):291-307.
James, William. 1890. *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
Jarvenpaa, S. L., and D. E. Leidner. 1999. "Communication and Trust in Global Virtual Teams." *Organization Science* 10(6):791-815.
Jensen, Michael, and William Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure." *Journal of Financial Economics* 3(4):305-360.
Johansson-Stenman, OIof, Mahmud Minhaj, and Peter Martinsson. 2005. "Does Stake Size Matter in Trust Games?" *Economics Letters* 88(3):365-369.
Johnson-George, Cynthia, and Walter C. Swap. 1982. "Measurement of Specific Interpersonal Trust: Construction and Validation of a Scale to Assess Trust in a Specific Other." *Journal of Personality and Social Psychology* 43(6):1306-1317.
Johnson, Martin. 2003. "Timepieces: Components of Survey Question Response Latencies." *Political Psychology* 25(5):679-702.
Johnson, Noel D., and Alexandra A. Mislin. 2011. "Trust Games: A Meta-Analysis." *Journal of Economic Psychology* 32(5):865-889.
Jones, E. E., and T. S. Pittman. 1982. "Toward a General Theory of Strategic Self-Presentation." Pp. 231-262 in *Psychological Perspectives on the Self*, edited by J. Suls. Hillsdale, NJ: Erlbaum Publishers.
Jones, Gareth R., and Jennifer M. George. 1998. "The Experience and Evolution of Trust: Implications for Cooperation and Teamwork." *The Academy of Management Review* 23(3):531-546.
Jones, Karen. 1996. "Trust as an Affective Attitude." *Ethics* 107(1):4-25.
Judd, Charles M., James W. Downing, Roger A. Drake, and Jon A. Krosnick. 1991. "Some Dynamic Properties of Attitude Structures: Context-Induced Response Facilitation and Polarization." *Journal of Personality and Social Psychology* 60(2):193-202.
Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
Kahneman, Daniel. 2003. "A Perspective on Judgment and Choice." *American Psychologist* 58(9):697-720.
Kahneman, Daniel, and Shane Frederick. 2002. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment." Pp. 49-81 in *Heuristics and Biases: The Psychology of Intuitive Thought*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman. New York: Cambridge University Press.
Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decisions Under Risk." *Econometrica* 47(2):263-292.
Kandel, Eugene, and Edward P. Lazear. 1992. "Peer Pressure and Partnerships." *Journal of Political Economy* 100(4):801-817.
Kassebaum, Ulf Bernd. 2004. "Interpersonelles Vertrauen. Entwicklung eines Inventars zur Erfassung spezifischer Aspekte des Konstrukts." Hamburg: University of Hamburg.
Kay, Aaron C., and Lee Ross. 2003. "The Perceptual Push: The Interplay of Implicit Cues and Explicit Situational Construals on Behavioral Intentions in the Prisoner´s Dilemma." *Journal of Experimental Social Psychology* 39(6):634-643.
Kay, Aaron C., Christian S. Wheeler, John A. Bargh, and Lee Ross. 2004. "Material Priming: The Influence of Mundane Physical Objects on Situational Construal and Competitive Behavioral Choice." *Organizational Behavior and Human Decision Processes* 95(1):83-96.
Kay, Aaron C., Christian S. Wheeler, and Dirk Smeesters. 2008. "The Situated Person: Effects of Construct Accessability on Situation Construals and Interpersonal Perception." *Journal of Experimental Social Psychology* 44(2):275-291.
Keller, Johannes, Gerd Bohner, and Hans-Peter Erb. 2000. "Intuitive und heuristische Urteilsbildung - verschiedene Prozesse?" *Zeitschrift für Sozialspychologie* 31(2):87-101.
Kelley, Harold H. 1992. "Common-Sense Psychology and Scientific Psychology." *Annual Review of Psychology* 43:1-23.
Keren, Gideon. 2007. "Framing, Intentions, and Trust-Choice Incompatibility." *Organizational Behavior and Human Decision Processes* 103(2):238-255.
Keyton, Joann, and Faye L. Smith. 2009. "Distrust in Leaders: Dimensions, Patterns and Emotional Intensity." *Journal of Leadership & Organizational Studies* 16(1):6-18.

Kirchgässner, Gebhard. 2008. *Homo Oeconomicus. The Economic Model of Behavior and its Applications in Economics and other Social Sciences*, 3 ed. New York: Springer.
Knack, S., and P. Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics* 112(4):1251-1288.
Knight, Frank H. 1965. *Risk, Uncertainty and Profit*. New York: Harper & Row.
Kocher, Martin G., Peter Martinsson, and Martine Visser. 2008. "Does Stake Size Matter for Cooperation and Punishment?" *Economics Letters* 99(3):508-511.
Kolm, Serge-Christpohe, and Jean Mercier Ythier (eds.). 2006. *Handbook of the Economics of Giving, Altruism and Reciprocity. Volume 1: Foundations*. Elsevier.
Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr. 2005. "Oxytocin Increases Trust in Humans." *Nature* 435(2):673-676.
Krajbich, I., R. Adolphs, D. Tranel, N. L. Denburg, and Colin Camerer. 2009. "Economic Games Quantify Diminished Sense of Guilt in Patients with damage to the Prefrontal Cortex." *Journal of Neuroscience* 29(7):2188-2192.
Kramer, R. M., and Roy J. Lewicki. 2010. "Repairing and Enhancing Trust: Approaches to Reducing Organizational Trust Deficits." *Academy of Management Annals* 4(1):245-277.
Kramer, Roderick M. 1996. "Divergent Realities and Convergent Disappointments in the Hierarchic Relation: Trust and the Intuitive Auditor Model at Work." Pp. 216-245 in *Trust in Organizations*, edited by Roderick M. Kramer, and Tom R. Tyler. Thousand Oaks: Sage.
—. 1999. "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions." *Annual Review of Psychology* 50:569-598.
—. 2004. "Collective Paranoia: Distrust Between Social Groups." Pp. 136-166 in *Distrust*, edited by Russell Hardin. New York: Russel Sage Foundation.
—. 2006. "Trust as Situated Cognition: An Ecological Perspective on Trust Decisions." Pp. 68-83 in *Handbook of Trust Research*, edited by Reinhard Bachmann, and Akbar Zaheer. Cheltenham, UK; Northampton,USA: Edward Elgar Publishing.
Kramer, Roderick M., and Karen S. Cook. 2004. *Trust and Distrust in Organizations. Dilemmas and Approaches*. New York: Russel Sage.
Kreps, David M. 1990. "Corporate Culture and Economic Theory." Pp. 90-143 in *Perspectives on Positive Political Economy*, edited by James E. Alt, and Kenneth A. Shepsle. Cambridge: Cambridge University Press.
Kreps, David M., and Robert Wilson. 1982. "Reptutation and Imperfect Information." *Journal of Economic Theory* 27(2):253-279.
Kromrey, Jeffrey M., and Lynne Foster-Johnson. 1998. "Mean Centering in Moderated Multiple Regression: Much Ado About Nothing." *Educational and Psychological Measurement* 58(1):42-67.
Kroneberg, C., I. Heintze, and G. Mehlkop. 2010a. "The Interplay of Moral Norms and Instrumental Incentives in Crime Causation." *Criminology* 48(1):259-294.
Kroneberg, C., M. Yaish, and V. Stocké. 2010b. "Norms and Rationality in Electoral Participation and the Rescue of Jews in WWII." *Rationality and Society* 22(1):3-36.
Kroneberg, Clemens. 2006a. "The Definition of the Situation and Variable Rationality: The Model of Frame Selection as a General Theory of Action." in *Working Paper Series, No. 06-05*. Sonderforschungsbereich 504, Universität Mannheim.
—. 2006b. "Die Erklärung der Wahlteilnahme und die Grenzen des Rational-Choice-Ansatzes. Eine Anwendung des Modells der Frame-Selektion." Pp. 79-111 in *Jahrbuch für Handlungs- und Entscheidungstheorie, Band 4, Schwerpunktthema Wahlen*, edited by J. Behnke, and T. Bräuninger. Wiesbaden: VS Verlag.
—. 2011a. *Die Erklärung sozialen Handelns. Grundlagen und Anwendungen einer integrativen Theorie*. Wiesbaden: VS Verlag.
—. 2011b. "Zusatzkapitel 1 zu 'Die Erklärung sozialen Handelns': Die Ableitung von Dreifach-Interaktionshypothesen aus dem Modell der Frame-Selektion." URL: http://vs-verlag.de/tu/Kroneberg-Erklaerung (Last Accessed 02/07/2012).
—. 2011c. "Zusatzkapitel 2 zu 'Die Erklärung sozialen Handelns': Statistische Modellierung und Testbarkeit des Modells." URL: http://vs-verlag.de/tu/Kroneberg-Erklaerung (Last Accessed 02/07/2012).
Krosnick, J. A., and R. P. Abelson. 1992. "The Case for Measuring Attitude Strength in Surveys." Pp. 177-203 in *Questions About Survey Questions*, edited by J. Tanur. New York: Russel Sage Foundation.
Krosnick, J. A., and R. E. Petty. 1995. "Attitude Strength: An Overview." Pp. 1-24 in *Attitude Strength: Antecedents and Consequences*, edited by R. E. Petty, and J. A. Krosnick. Hillsdale, NJ: Lawrence Erlbaum.
Krosnick, Jon A., David S. Boninger, Yao C. Chuang, Matthew K. Berent, and Catherine G. Carnot. 1993. "Attitude Strength: One Construct or Many Related Constructs?" *Journal of Personality and Social Psychology* 65(6):1132-1151.

Krueger, Frank, Kevin McCabe, Jorge Moll, Nikolaus Kriegeskorte, Roland Zahn, Maren Strenziok, Armin Heinecke, and Jordan Grafman. 2007. "Neural Correlates of Trust." *PNAS* 104(50):20084-20089.
Kruglanski, A., and E. P. Thompson. 1999. "Persuasion by a Single Route: A View From the Unimodel." *Psychological Inquiry* 10(2):83-109.
Kugler, Tamar, Terry Connolly, and Edgar E. Kausel. 2009. "The Effect of Consequential Thinking on Trust Game Behavior." *Journal of Behavioral Decision Making* 22(2):101-119.
Kuhberger, Anton. 1998. "The Influence of Framing on Risky Decisions: A Meta-Analysis." *Organizational Behavior and Human Decision Processes* 75(1):23-55.
La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and Vishny R. 1997. "Trust in Large Organisations." *American Economic Review* 87(2):333-338.
Lagerspetz, Olli. 2001. "Vertrauen als geistiges Phänomen." Pp. 85-113 in *Vertrauen - Die Grundlage des sozialen Zusammenhalts*, edited by Martin Hartmann, and Claus Offe. Frankfurt, New York: Campus.
Lahno, Bernd. 2001. "On the Emotional Character of Trust." *Ethical Theory and Moral Practice* 4:171-189.
—. 2002. *Der Begriff des Vertrauens*. Paderborn: mentis.
Lance, Charles E. 1988. "Residual Centering, Exploratory and Confirmatory Moderator Analysis, and Decomposition of Effects in Path Models Containing Interactions." *Applied Psychological Measurement* 12(2):163-175.
Langlois, Judith H., Lisa Klakanis, Adam Rubenstein, Andrea Larson, Monica Hallam, and Monica Smoot. 2000. "Maxims or Myths of Beauty? A Meta-Analysis and Theoretical Review." *Psychological Bulletin* 126(3):390-423.
Larsen, Jeff T., Peter A. McGraw, and John T. Cacioppo. 2001. "Can People Feel Happy and Sad at the Same Time?" *Journal of Personality and Social Psychology* 81(4):684-696.
Larzelere, Robert E., and Ted L. Huston. 1980. "The Dyadic Trust Scale: Toward Understanding Interpersonal Trust in Close Relationships." *Journal of Marriage and Family* 42(3):595-604.
Lavine, H., E. Borgida, and J. L. Sullivan. 2000. "On the Relationship Between Attitude Involvement and Attitude Accessibility: Toward a Cognitive-Motivational Model of Political Information Processing " *Political Psychology* 21(1):81-106.
Lazarus, Richard S. 1991a. "Cognition and Motivation in Emotion." *American Psychologist* 46(4):352-367.
—. 1991b. "Progress on a Cognitive-Motivational-Relational Theory of Emotion." *American Psychologist* 46(8):819-834.
Lazzarini, Sergio G., Regina Madalozzo, Rinaldo Artes, and José de Oliveira Siqueira. 2003. "Measuring Trust: An Experiment In Brazil." in *Insper Working Paper, No. 049/2004*. IBMEC Business School, Sâo Paulo.
Leary, Mark R., and Robin M. Kowalski. 1990. "Impression Management: A Literature Review and Two-Component Model." *Psychological Bulletin* 107(1):34-47.
Lenton, Pamela, and Paul Mosley. 2011. "Incentivising Trust." *Journal of Economic Psychology* 32(5):890-897.
Levin, I. P., D. P. Chapman, and R. D. Johnson. 1988. "Confidence in Judgments Based on Incomplete Information: An Investigation Using Both Hypothetical and Real Gambles." *Journal of Behavioral Decision Making* 1(1):29-41.
Levin, Irwin P., Sandra L. Schneider, and Gary J. Gaeth. 1998. "All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects." *Organizational Behavior and Human Decision Processes* 76(2):149-188.
Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1(3):593-622.
Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21(2):153-174.
Lewicki, Roy J. 2006. "Trust and Distrust." Pp. 191-203 in *The Negotiator´s Fieldbook*, edited by Andrea Kupfer Schneider, and Christopher Honeyman. Washington: American Bar Association.
Lewicki, Roy J., and Barbara B. Bunker. 1995a. "Developing and Maintaining Trust in Work Relationships." Pp. 114-139 in *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer, and Tom R. Tyler. Thousand Oaks, CA: Sage.
—. 1995b. "Trust in Relationships: A Model of Development and Decline." Pp. 133-173 in *Conflict, Cooperation, and Justice: Essays Inspired by the Work of Morton Deutsch*, edited by Barbara B. Bunker, and Jeffrey Z. Rubin. San-Francisco: Jossey-Bass.
—. 1996. "Developing and Maintaining Trust in Work Relationships." Pp. 114-139 in *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer, and Tom R. Tyler. Thousand Oaks, CA: Sage.
Lewicki, Roy J., Daniel J. McAllister, and Robert J. Bies. 1998. "Trust and Distrust: New Relationships and Realities." *Academy of Management Review* 23(3):438-458.

Lewicki, Roy J., Edward C. Tomlinson, and Nicole Gillespie. 2006. "Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions." *Journal of Management* 32(6):991-1022.
Lewis, J. David, and Andrew J. Weigert. 1985a. "Social Atomism, Holism, and Trust." *The Sociological Quarterly* 26(4):455-471.
—. 1985b. "Trust as a Social Reality." *Social Forces* 63(4):967-985.
Liberman, Matthew D. 2006. "Social Cognitive Neuroscience: A Review of Core Processes." *Annual Review of Psychology* 58:259-289.
Liberman, Varda, Steven M. Samuels, and Lee Ross. 2004. "The Name of the Game: Predictive Power of Reputations versus Situational Labels in Determining Prisoner´s Dilemma Game Moves." *Personality and Social Psychology Bulletin* 30(9):1175-1185.
Lichtenstein, S., and P. Slovic. 1971. "Reversals of Preferences Between Bids and Choices in Gambling Decisions." *Journal of Experimental Psychology* 89(1):46-55.
Lieberman, David A. 2007. "Social Cognitive Neuroscience: A Review of Core Processes." *Annual Review of Psychology* 58:259-289.
Lindenberg, Siegwart. 1989. "Social Production Functions, Deficits, and Social Revolutions: Prerevolutionary France and Russia." *Rationality and Society* 1(1):51-77.
—. 1992. "The Method of Decreasing Abstraction." Pp. 3-20 in *Rational Choice Theory. Advocacy and Critique*, edited by James S. Coleman, and Thomas J. Fararo. Newbury Park (a.o.): Sage.
—. 2000. "It Takes Both Trust and Lack of Mistrust: The Workings of Cooperation and Relational Signaling in Contractual Relationships." *Journal of Management and Governance* 4(1):11-33.
—. 2003. "Governance Seen From a Framing Point of View: The Employment Relationship and Relationship Signalling." Pp. 37-57 in *The Trust Process in Organizations*, edited by Bart Nooteboom, and Frédérique Six. Cheltenham: Edward Elgar Publishing.
List, J. A., and T. L. Cherry. 2008. "Examining the Role of Fairness in High Stakes Allocation Decisions." *Journal of Economic Behavior & Organization* 65(1):1-8.
Little, Todd D., James A. Bovaird, and Keith F. Widaman. 2006. "On the Merits of Orthogonalizing Powered and Product Terms: Implications for Modeling Interactions Among Latent Variables." *Structural Equation Modeling* 13(4):497-519.
Lohrenz, Terry, Kevin McCabe, Colin F. Camerer, and P. Read Montague. 2007. "Neural Signature of Fictive Learning Signals in a Sequential Investment Task." *PNAS* 104(22):9493-9498.
Loomis, James L. 1959. "Communication, the Development of Trust, and Cooperative Behavior." *Human Relations* 12:305-315.
Lorenz, Edward. 1999. "Trust, Contract and Economic Cooperation." *Cambridge Journal of Economics* 23(3):301-315.
Lount, Robert B. Jr. 2010. "The Impact of Positive Mood on Trust in Interpersonal and Intergroup Interactions." *Journal of Personality and Social Psychology* 98(3):420-433.
Luce, Duncan R. 1986. *Response Times. Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.
Luhmann, Niklas. 1979. *Trust and Power*. Chichester (a.o.): John Wiley & Sons.
—. 1988. "Familiarity, Confidence, Trust: Problems and Alternatives." Pp. 94-107 in *Trust: Making and Braking Cooperative Relations*, edited by Diego Gambetta. New York: Blackwell.
—. 1990. "The Autpoiesis of Social Systems." Pp. 1-20 in *Essays on Self-Reference*, edited by Niklas Luhmann. New York: Columbia University Press.
—. 1995. *Social Systems*. Stanford: Stanford University Press.
—. 2000. *Vertrauen - Ein Mechanismus zur Reduktion sozialer Komplexität*, 4 ed. Stuttgart: Lucius & Lucius.
Macrae, Neil C., and Galen V. Bodenhausen. 2000. "Social Cognition: Thinking Categorially about Others." *Annual Review of Psychology* 51:93-120.
Maddala, G. S. 1991. "A Perspective on the Use of Limited-Dependent and Qualitative Variables Models in Accounting Research." *The Accounting Review* 66(4):788-807.
Maier, Victoria Elizabeth. 2009. "The Role of Emotion in Leader Trust Processes." St. Gallen: University of St. Gallen.
Main, M., N. Kaplan, and J. Cassidy. 1985. "Security in Infancy, Childhood, and Adulthood: A Move to the Level of Presentation." *Monographs of the Society for Research in Child Development* 50(1/2):66-104.
Malhotra, Deepak. 2004. "Trust and Reciprocity Decisions: The Differing Perspectives of Trustors and Trusted Parties." *Organizational Behavior and Human Decision Processes* 94(2):61-73.
Malhotra, Deepak, and J. Keith Murnighan. 2002. "The Effects of Contracts on Interpersonal Trust." *Administrative Science Quarterly* 47(3):534-559.
March, James G. 1978. "Bounded Rationality, Ambiguity, and the Engineering of Choice." *The Bell Journal of Economics* 9(2):587-608.

—. 1994. *A Primer on Decision Making*. New York: Free Press.
March, James G., and Johan P. Olsen. 1989. *Rediscovering Institutions: The Organizational Basis of Politics*. New York: Free Press.
Markus, Hazel. 1977. "Self-Schemata and Processing Information About the Self." *Journal of Personality and Social Psychology* 35(2):63-78.
Markus, Hazel, and Elissa Wurf. 1987. "The Dynamic Self Concept: A Social Psychological Perspective." *Annual Review of Psychology* 38:299-337.
Marschak, Jacob. 1975. "Personal Probabilites of Probabilites." *Theory and Decision* 6(2):121-153.
Mas-Colell, Andreu, Micheal D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. New York, Oxford: Oxford University Press.
Matzke, Dora, and Eric-jan Wagenmakers. 2009. "Psychological Interpretation of the ex-Gaussian and Shifted Wald Parameters: A Diffusion Model Analysis." *Psychonomic Bulletin & Review* 16(5):798-817.
Mayer, Roger C., James H. Davis, and David F. Schoorman. 1995. "An Integrative Model of Organizational Trust." *The Academy of Management Review* 20(3):709-734.
Mayerl, Jochen. 2009. *Kognitive Grundlagen sozialen Verhaltens. Framing, Einstellungen und Rationalität*. Wiesbaden: VS Verlag.
—. 2010. "Die Low-Cost-Hypothese ist nicht genug." *Zeitschrift für Soziologie* 39(1):38-59.
Mayerl, Jochen, and Dieter Urban. 2008. *Antwortreaktionszeiten in Survey-Analysen. Messung, Auswertung und Anwendung*. Wiesbaden: VS Verlag.
McAllister, Daniel J. . 1995. "Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations." *Academy of Management Journal* 38(1):24-59.
McCabe, Kevin, Daniel Houser, Lee Ryan, Vernon L. Smith, and Theodore Trouard. 2001. "A Functional Imaging Study of Cooperatino in Two-Person Reciprocal Exchange." *PNAS* 98(20):11898-11895.
McEvily, Bill, Vincenzo Perrone, and Akbar Zaheer. 2003. "Trust as an Organizing Principle." *Organization Science* 14(1):91-103.
McKnight, Harrison D., and Norman L. Chervany. 1996. "The Meanings of Trust." in *Carlson School of Management Working Paper, No. 96-04*. University of Minnesota.
—. 2000. "What is Trust? A Conceptual Analysis and an Interdisciplinary Model." in *AMCIS 2000 Proceedings, Paper 382*. URL: http://aisel.aisnet.org/amcis2000/382 (Last Accessed: 11/04/2011).
—. 2001. "Trust and Distrust Definitions: One Bite at a Time." Pp. 27-54 in *Trust in Cyber-Societies*, edited by R. Falcone, M. Singh, and Y.H. Tan. Berlin, Heidelberg: Springer.
—. 2006. "Reflections on an Intitial Trust-Building Model." Pp. 29-51 in *Handbook of Trust Research*, edited by Reinhard Bachmann, and Akbar Zaheer. Cheltenham, UK; Northampton,USA: Edward Elgar Publishing.
McKnight, Harrison D., Larry L. Cummings, and Norman L. Chervany. 1998. "Intitial Trust Formation in New Organizational Relationships." *Academy of Management Review* 23(3):473-490.
Mead, George H. 1934. *Mind, Self and Society*. Chicago: University of Chicago Press.
—. 1967. *Mind, Self, and Society. From the Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.
Mellers, B.A., A. Schwartz, and A.D.J. Cooke. 1998. "Judgment and Decision Making." *Annual Review of Psychology* 49:447-477.
Messick, David M. 1999. "Alternative Logics for Decision Making in Social Settings." *Journal of Economic Behavior & Organization* 39(1):11-28.
Messick, David M., and Roderick M. Kramer. 2001. "Trust as a Shallow Form of Morality." Pp. 89-117 in *Trust in Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Messick, David M., and Diane M. Mackie. 1989. "Intergroup Relations." *Annual Review of Psychology* 40:45-81.
Meyerson, Debra, Karl E. Weick, and Roderick M. Kramer. 1996. "Swift Trust and Temporary Groups." Pp. 166-195 in *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer, and Tom R. Tyler. Thousand Oaks, CA: Sage.
Mikulincer, Mario. 1998. "Attachment Working Models and the Sense of Trust: An Exploration of Interaction Goald and Affect Regulation." *Journal of Personality and Social Psychology* 74(5):1209-1224.
Millar, F. E., and L. E. Rogers. 1976. "A Relational Approach to Interpersonal Communication." Pp. 87-104 in *Explorations in Interpersonal Communication*, edited by G. R. Miller. Beverly Hills, CA: Sage.
Miller, Garry. 2001. "Why is Trust necessary in Organizations? The Moral Hazard of Profit Maximization." Pp. 307-331 in *Trust in Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Mills, J., and M. C. Clark. 1994. "Communal and Exchange Relationships: Controversies and Research." Pp. 29-42 in *Theoretical Frameworks for Personal Relationships*, edited by Raplh Erber, and Robin Gilmour. Hillsdale: Lawrence Erlbaum Associates.

Mischel, Walter. 1977. "The Interaction of Person and Situation." Pp. 333-352 in *Personality at the Crossroads: Current Issues in Interactional Psychology*, edited by D. Magnusson, and N. S. Endler. Hillsdale, NJ: Erlbaum Associates.
Mischel, Walter, and Yuichi Shoda. 1995. "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics and Invariance in Personality Structure." *Psychological Review* 102(2):246-268.
Mishra, A. K. 1996. "Organizational Responses to Crisis: The Centrality of Trust." Pp. 261-287 in *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer, and Tom R. Tyler. Thousand Oaks, CA: Sage.
Misztal, Barbara A. 1996. *Trust in Modern Societies*. Cambridge: Polity Press.
—. 2001. "Normality and Trust in Goffman´s Theory of Interaction Order." *Sociological Theory* 19(3):312-324.
Mitchell, S.A. 1988. *Relational Concepts in Psychoanalysis*. Cambridge, MA: Harvard University Press.
Molden, D. C:, and E. T. Higgins. 2005. "Motivated Thinking." Pp. 295-317 in *The Cambridge Handbook of Thinking and Reasoning*, edited by K. Holyoak, and B. Morrison. New York: Guilford Press.
Möllering, Guido. 2001. "The Nature of Trust: From Georg Simmel to a Theory of Expectation, Interpretation and Suspension." *Sociology* 35(2):403-420.
—. 2005a. *Trust Under Pressure: Empirical Investigations of Trust and Trust Building in Uncertain Circumstances*. Cheltenham: Edward Elgar Publishing.
—. 2005b. "The Trust/Control Duality: An Integrative Perspective on Positive Expectations of Others." *International Sociology* 20(3):283-305.
—. 2006a. "Trust, Institutions, Agency: Towards a Neoinstitutional Theory of Trust." Pp. 355-376 in *Handbook of Trust Research*, edited by Reinhard Bachmann, and Akbar Zaheer. Cheltenham, UK; Northampton,USA: Edward Elgar Publishing.
—. 2006b. *Trust: Reason, Routine, Reflexivity*. Amsterdam (a.o.): Elsevier.
Mulder, G. 1986. "The Concept and Measurement of Mental Effort." Pp. 175-198 in *Energetics and Human Information Processing*, edited by Robert J. Hockey, Anthony W. K. Gaillard, and Michauel G. H. Coles. Dordrecht: Kluwer.
Mulder, Laetitia B. 2008. "The Difference Between Punishments and Rewards in Fostering Moral Concerns in Social Decision Making." *Journal of Experimental Social Psychology* 44(6):1436-1443.
Mulder, Laetitia B., Eric van Dijk, David De Cremer, and Henk A.M. Wilke. 2006. "Undermining Trust and Cooperation: The Paradox of Sanctioning Systems in Social Dilemmas." *Journal of Experimental Social Psychology* 42(2):147-162.
Müller, Walter, Susanne Steinmann, and Renate Ell. 1998. "Education and Labour-Market Entry in Germany." Pp. 143-188 in *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations*, edited by Walter Müller, and Yossi Shavit. Oxford: Clarendon Press.
Mulligan, K., J. T. Grant, S. T. Mockabee, and J. Q. Monson. 2003. "Response Latency Methodology for Survey Research: Measurement and Modeling Strategies." *Political Analysis* 11(3):289-301.
Naef, Michael, and Jürgen Schupp. 2009. "Measuring Trust: Experiments and Surveys in Contrast and Combination." in *SOEP Papers on Multidisciplinary Panel Data Research, No. 167*. Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.
Nash, John F. 1950. "Equilibrium in N-Player Games." *Proceedings of the National Academy of the Sciences (PNAS)* 36(1):48-49.
Nauck, Bernhard. 2010. "Fertilitätsstrategien im interkulturellen Vergleich: Value of Children, ideale und angestrebte Kinderzahl in zwölf Ländern." Pp. 213-238 in *Psychologie - Kultur - Gesellschaft*, edited by Boris Mayer, and Hans-Joachim Kornadt. Wiesbaden: VS Verlag für Sozialwissenschaften.
Nooteboom, Bart. 2002. *Trust: Forms, Foundations, Functions, Failures and Figures*. Cheltenham: Edward Elgar.
—. 2006. "Forms, Sources and Processes of Trust." Pp. 247-263 in *Handbook of Trust Research*, edited by Reinhard Bachmann, and Akbar Zaheer. Cheltenham, UK; Northampton,USA: Edward Elgar Publishing.
—. 2007. "Methodological Interactionism: Theory and Application to the Firm and to the Building of Trust." *Review of Austrian Economics* 20:137-153.
Olson, James M., and Mark P. Zanna. 1993. "Attitudes and Attitude Change." *Annual Review of Psychology* 44:117-154.
Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
Opp, Karl-Dieter. 1999. "Contending Conceptions of the Theory of Rational Action." *Journal of Theoretical Politics* 11(2):171-202.
—. 2010. "Frame-Selektion, Normen und Rationalität." *Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft* 50:63-78.

Orbell, John, Robyn Dawes, and Peregrine Schwartz-Shea. 1994. "Trust, Social Categories, and Individuals: The Case of Gender." *Motivation and Emotion* 18(2):109-128.
Orbell, John M., Alphons J. C. van de Kragt, and Robyn M. Dawes. 1988. "Explaining Discussion-Induced Cooperation." *Journal of Personality and Social Psychology* 54(5):811-819.
Ostrom, Elinor. 1998. "A Behavioural Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997." *American Political Science Review* 92:1-22.
—. 2000. "Collective Action and the Evolution of Social Norms." *Journal of Economic Perspectives* 14(1):137-158.
—. 2003. "Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation." Pp. 19-79 in *Trust and Reciprocity. Interdisciplinary Lessons from Experimental Research*, edited by Elinor Ostrom, and James Walker. New York: Russel Sage Foundation.
Ostrom, Elinor, and James Walker. 2003. *Trust and Reciprocity - Interdisciplinary Lessons from Experimental Research*. New York: Russel Sage Foundation.
Pacini, Rosemary, and Seymour Epstein. 1999. "The Relation of Rational and Experiential Information Processing Styles to Personality, Basic Beliefs, and the Ratio-Bias Phenomenon." *Journal of Personality and Social Psychology* 76(6):972-987.
Papke, Leslie E., and Jeffrey M. Wooldrige. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates." *Journal of Applied Econometrics* 11(6):619-632.
Parsons, Talcott. 1963. "On the Concept of Influence." *Public Opinion Quarterly* 27(1):37-62.
—. 1967. *Sociological Theory and Modern Society*. New York: Free Press.
—. 1971. *The System of Modern Societies*. Englewood Cliffs, NJ: Prentice-Hall.
—. 1978. "Research Within Human Subjects and the 'Professional Complex'." Pp. 264-296 in *Action Theory and the Human Condition*, edited by Talcott Parsons. New York, London: Free Press.
Payne, John W. 1976. "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis." *Organizational Behavior and Human Decision Processes* 16(2):366-387.
Payne, John W., James R. Bettman, and Eric J. Johnson. 1988. "Adaptive Strategy Selection in Decision Making." *Journal of Experimental Psychology: Learning, Memory and Cognition* 14(3):534-552.
—. 1992. "Behavioral Decision Research: A Constructive Processing Perspective." *Annual Review of Psychology* 43:87-131.
—. 1993. *The Adaptive Decision Maker*. Cambridge, NY: Cambridge University Press.
Perugini, Marco, Marcello Gallucci, Fabio Presaghi, and Anna Paola Ercolani. 2003. "The Personal Norm of Reciprocity." *European Journal of Personality* 17(4):251-283.
Petty, R. E., and J.T. Cacioppo. 1986. "The Elaboration Likelihood Model of Persuasion." *Advances in Experimental Social Psychology* 19:124-205.
Petty, R. E., and D. T. Wegener. 1999. "The Elaboration Likelihood Model:: Current Status and Controversies." Pp. 41-72 in *Dual-Process Theories in Social Psychology*, edited by Shelly Chaiken, and Yaacov Trope. New York, London: Guilford Press.
Phelps, Elizabeth A. 2006. "Emotion and Cognition: Insights from Studies of the Human Amygdala." *Annual Review of Psychology* 57:27-53.
Pietromonaco, Paula R., and Lisa F. Barrett. 2000. "The Internal Working Model Concept: What Do We Really Know About the Self in Relation to Others?" *Review of General Psychology* 4(2):155-175.
Planalp, S. 1987. "Interplay Between Relational Knowledge and Events." Pp. 175-191 in *Accounting for Relationships*, edited by R. Burnett, P. McGhee, and D. D. Clarke. New York: Methuen.
Pomerantz, E. M., S. Chaiken, and R. S. Tordesillas. 1995. "Attitude Strength and Resistance Processes." *Journal of Personality and Social Psychology* 69(3):408-419.
Popper, Karl R. 1945. *The Open Society and Its Enemies*. London: Routledge & Keagan Paul.
Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24:1-24.
Priester, Joseph R., and Richard E. Petty. 1996. "The Gradual Threshold Model of Ambivalence: Relating the Positive and Negative Bases of Attitudes to Subjective Ambivalence." *Journal of Personality and Social Psychology* 71(3):431-449.
Putnam, Robert D. 1993. *Making Democracy Work. Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
—. 1995. "Bowling Alone: America´s Declining Social Capital." *Journal of Democracy* 6(1):65-75.
Rabin, Matthew. 1993. "Incorporating Fairness Into Game Theory." *American Economic Review* 83(5):1281-1302.
—. 1998. "Psychology and Economics." *Journal of Economic Literature* 36(1):11-46.

Rapoport, Amnon, and Albert M. Chammah. 1965. *Prisoner´s Dilemma: A Study in Conflict an Cooperation*. Ann Arbor: University of Michigan Press.
Ratcliff, Roger. 1978. "A Theory of Memory Retrieval." *Psychological Review* 85(2):59-108.
—. 1993. "Methods for Dealing With Reaction Time Outliers." *Psychological Bulletin* 114(3):510-532.
Ratcliff, Roger, Trisha Van Zandt, and Gail McKoon. 1999. "Connectionist and Diffusion Models of Reaction Time." *Psychological Review* 106(2):261-300.
Raub, Werner. 2004. "Hostage Posting as a Mechanism of Trust: Binding, Compensation, and Signaling." *Rationality and Society* 16(3):319-365.
Raub, Werner, and Jeroen Weesie. 1990. "Reputation and Efficiency in Social Interactions: An Example of Network Effects." *American Journal of Sociology* 96(3):626-654.
Rauhut, Heiko, and Ivar Krumpal. 2008. "Die Durchsetzung sozialer Normen in Low-Cost und High-Cost-Situationen." *Zeitschrift für Soziologie* 37(5):380-402.
Reber, R., and N. Schwarz. 1999. "Effects of Perceptual Fluency on Judgments of Truth." *Consciousness and Cognition* 8(3):338-342.
Reber, R., N. Schwarz, and P. Winkielman. 2004. "Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver´s Processing Experience?" *Personality and Social Psychology Review* 8(4):364-382.
Reis, Harry T., Andrew W. Collins, and Ellen Berscheid. 2000. "The Relationship Context of Human Behavior and Development." *Psychological Bulletin* 126(6):844-872.
Rempel, J. K., J. G. Holmes, and M.P. Zanna. 1985. "Trust in Close Relationships." *Journal of Personality and Social Psychology* 49(1):95-112.
Rempel, J. K., M. Ross, and J. G. Holmes. 2001. "Trust and Communicated Attributions in Close Relationships." *Journal of Personality and Social Psychology* 81(1):57-64.
Ren, Hong, and Barbara Gray. 2009. "Repairing Relationship Conflict: How Violation Types and Culture Influence the Effectiveness of Restoration Rituals." *Academy of Management Review* 34(1):105-126.
Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2009. "Is Mistrust Self-Fulfilling?" *Economics Letters* 104(2):89-91.
Rice, Ronald E. 1992. "Task Analyzability, Use of New Media and Effectiveness: A Multi-Site Exploration of Media Richness." *Organization Science* 3(4):475-500.
Riker, William H., and Peter C. Ordeshook. 1973. *An Introduction to Positive Political Theory*. Englewood-Cliffs, NJ: Prentice-Hall.
Rilling, James K. , David A. Gutman, Thorsten R. Zeh, Giuseppe Pagnoni, Gregory S. Berns, and Clinton D. Kilts. 2002. "A Neural Basis for Social Cooperation." *Neuron* 35:395-405.
Rilling, James K., and Alan G. Sanfey. 2011. "The Neuroscience of Social Decision Making." *Annual Review of Psychology* 62:23-48.
Ripperger, Tanja. 1998. *Ökonomik des Vertrauens. Analyse eines Organisationsprinzips*. Tübingen: Mohr Siebeck.
Robinson, Sandra L. 1996. "Trust and Breach of the Psychological Contract." *Administrative Science Quarterly* 41(4):574-599.
Robinson, Sandra L., and Denise M. Rousseau. 1994. "Violating the Psychological Contract: Not the Exception, but the Norm." *Journal of Organizational Behavior* 15(3):245-259.
Rockmann, Kevin W., and Gregory B. Northcraft. 2008. "To Be or not to Be Trusted: The Influence of Media Richness on Defection and Deception." *Organizational Behavior and Human Decision Processes* 107(2):106-122.
Rokeach, Milton. 1968. "A Theory of Organization and Change Within Value-Attitude Systems." *Journal of Social Issues* 24(1):13-33.
—. 1973. *The Nature of Human Values*. New York: Free Press.
Rompf, Stephan A. 2008. "Vertrauen und Kommunikation. Variable Rationalität und die Frame-Perspektive des Vertrauens." Unpublished Thesis. Mannhein: University of Mannheim.
Ross, Lee, David Greene, and Pamela House. 1977. "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes." *Journal of Experimental Social Psychology* 13(3):279-301.
Ross, Lee, and Andrew Ward. 1996. "Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding." Pp. 103-135 in *Values and Knowledge*, edited by Edward S. Reed, Elliot Turiel, and Terrence Brown. Mawah, NJ: Lawrence Erlbaum Associates.
Rotter, J. B. 1967. "A New Scale for the Measurement of Interpersonal Trust." *Journal of Personality* 35(4):615-665.
—. 1971. "Generalized Expectancies for Interpersonal Trust." *American Psychologist* 26(5):443-452.
—. 1980. "Interpersonal Trust, Trustworthiness, and Gullibility." *American Psychologist* 35(1):1-7.
Rouder, Jeffrey N., Dngchu Sun, Paul L. Speckman, Jun Lu, and Duo Zhou. 2003. "A Hierarchical Bayesian Statistical Framework for Response Time Distributions." *Psychometrika* 68(4):589-606.

Rousseau, Denise M. 1989. "Psychological and Implied Contracts in Organizations." *Employee Responsibilities and Rights Journal* 2(2):121-139.
—. 1995. *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements.* Newbury Park, CA: Sage.
Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. "Not so Different after All: A Cross-Discipline View of Trust." *The Academy of Management Review* 23(3):303-404.
Rumelhart, David E. 1980. "Schemata: The Building Blocks of Social Cognition." Pp. 33-58 in *Theoretical Issues in Reading Comprehension*, edited by R. J. Spiro, B. C. Bruce, and W. E. Brewer. Hillsdale, NJ: Lawrence Erlbaum Associates.
Russell, James A., and James M. Carroll. 1999. "On the Bipolarity of Positive and Negative Affect." *Psychological Bulletin* 125(1):3-30.
Safran, J.D. 1990a. "Towards a Refinement of Cognitive Therapy in Light of Interpersonal Theory: I. Theory." *Clinical Psychology Review* 10(1):87-105.
—. 1990b. "Towards a Refinement of Cognitive Therapy in Light of Interpersonal Theory: II. Practice." *Clinical Psychology Review* 10(1):107-121.
Sally, David. 1995. "Conversation and Cooperation in Social Dilemmas. A Meta-Analysis of Experiments From 1958 to 1992." *Rationality and Society* 7(1):58-92.
Sanbonmatsu, David M., and Russell H. Fazio. 1990. "The Role of Attitudes in Memory-Based Decision Making." *Journal of Personality and Social Psychology* 59(4):614-622.
Sanfey, Alan G. 2007. "Social Decision-Making: Insights from Game Theory and Neuroscience." *Science* 318(5850):598-602.
Sapienza, Paola, Anna Toldra, and Luigi Zingales. 2008. "Understanding Trust." in *NBER Working Paper Series, No. 13387*. National Bureau of Economic Research, Cambridge, MA.
Saunders, Mark N., Denise Skinner, Graham Dietz, Nicole Gillespie, and Roy J. Lewicki. 2010. *Organizational Trust. A Cultural Perspective*. Cambridge (a.o.): Cambridge University Press.
Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Wiley.
Schank, R. C., and R. P. Abelson. 1977. *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Schechter, Laura. 2007. "Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay." *Journal of Economic Behavior & Organization* 62(2):272-292.
Scheuerer-Englisch, Hermann, and Peter Zimmerman. 1997. "Vertrauensentwicklung in Kindheit und Jugend." Pp. 27-48 in *Interpersonales Vertrauen. Theorien und Empirische Befunde*, edited by Martin Schweer. Opladen, Wiesbaden: Westdeutscher Verlag.
Schlenker, B. R. 1980. *Impression Management: The Self-Concept, Social Identity, and Interpersonal Relations*. Monterey, CA: Brooks/Cole.
Schoemaker, Paul J. H. 1982. "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations " *Journal of Economic Literature* 20(2):529-563.
Schoorman, David F., Roger C. Mayer, and James H. Davis. 2007. "An Integrative Model of Organizational Trust: Past, Present and Future." *Academy of Management Review* 32(2):344-354.
Schul, Yaacov, Ruth Mayo, and Eugene Burnstein. 2008. "The Value of Distrust." *Journal of Experimental Social Psychology* 44(5):1293-1302.
Schunk, Daniel, and Cornelia Betsch. 2006. "Explaining Heterogeneity Differences in Utility Functions by Individual Differences in Decision Modes." *Journal of Economic Psychology* 27(3):381-401.
Schütz, Alfred. 1967. *The Phenomenology of the Social World*. Evanston: Northwestern University Press.
Schütz, Alfred, and Thomas Luckmann. 1973. *The Structures of the Life World.* Evanston: Northwestern University Press.
Schwarz, N., F. Strack, H. Bless, G. Klumpp, H. Rittenauer-Schatka, and A. Simons. 1991. "Ease of Retrieval as Information: Another Look at the Availability Heuristic." *Journal of Personality and Social Psychology* 61(2):195-202.
Schwarz, Norbert. 1990. "Feelings as Information: Informational and Motivational Functions of Affective States." Pp. 527-561 in *Handbook of Motivation and Cognition: Foundations of Social Behavior, Vol. 2*, edited by E. T. Higgins, and R.M. Sorrentino. New York: Guilford Press.
—. 1998. "Warmer and More Social: Recent Developments in Cognitive Social Psychology." *Annual Review of Sociology* 24:239-264.
—. 2009. "Mental Construal in Social Judgment." Pp. 121-138 in *Social Cognition. The Basis of Human Interaction*, edited by Fritz Strack, and Jens Förster. New York, London: Psychology Press.
Schwarz, Norbert, and Gerald L. Clore. 1983. "Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States." *Journal of Personality and Social Psychology* 45(3):513.
—. 1996. "Feelings and Phenomenal Experiences." Pp. 433-465 in *Social Psychology: Handbook of Basic Principles*, edited by E. T. Higgins, and A. Kruglanski. New York, London: Guilford Press.

—. 2007. "Feelings and Phenomenal Experiences." Pp. 385-407 in *Social Psychology. A Handbook of Basic Principles*, edited by E. T. Higgins, and A. Kruglanski. New York: Guilford Press.
Seligman, Adam. 1997. *The Problem of Trust*. Princeton, NJ: Princeton University Press.
Selten, Reinhard. 1965. "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit." *Zeitschrift für die gesamte Staatswissenschaft* 121:301-324, 667-689.
—. 1975. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory* 4(1):25-55.
—. 2001. "What is Bounded Rationality." Pp. 13-36 in *Bounded Rationality: The Adaptive Toolbox*, edited by Gerd Gigerenzer, and Reinhard Selten. Cambridge, MA: MIT Press.
Servatka, Maros, Steven Tucker, and Radovan Vadovic. 2008. "Strategic Use of Trust." in *Working Paper Series, No. 11/2008*. Department of Economics, College of Business and Economics, University of Canterbury.
—. 2011. "Words Speak Louder Than Money." *Journal of Economic Psychology* 32(5):700-709.
Shapiro, Susan P. 1987. "The Social Control of Impersonal Trust." *American Journal of Sociology* 93(3):623-658.
Shaver, P. R., and C. Hazan. 1993. "Adult Romantic Attachment: Theory and Evidence." Pp. 29-70 in *Advances in Personal Relationships, Vol. 4*, edited by Daniel Perlman, and Warren H. Jones. London: Kingsley.
Sheppard, Blair H., and Dana M. Sherman. 1998. "The Grammars of Trust: A Model and General Implications." *Academy of Management Review* 23(3):422-437.
Shieh, Gwowen. 2011. "Clarifying the Role of Mean-Centering in Multicollinearity of Interaction Effects." *British Journal of Mathematical and Statistical Psychology* 64(3):462-477.
Shiloh, Shoshana, Efrat Salton, and Dana Sharabi. 2002. "Individual Differences in Rational and Intuitive Thinking Styles as Predictors of Heuristic Responses and Framing Effects." *Personality and Individual Differences* 32(3):415-429.
Shiv, B., and A. Fedorikhin. 2002. "Spontaneous Versus Controlled Influences of Stimulus-Based Affect on Choice Behavior." *Organizational Behavior and Human Decision Processes* 87(2):342-370.
Simmel, Georg. 1992. *Soziologie. Untersuchungen über die Formen der Vergesellschaftung*. Frankfut am Main: Suhrkamp.
Simon, Bernd. 2004. *Identity in Modern Society. A Social Psychological Perspective*. Malden, MA, USA; Oxford, UK; Victoria, Australia: Blackwell Publishing.
Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69(1):99-118.
—. 1978. "Rationality as Process and as Product of Thought." *American Economic Review* Papers and Proceedings of the 19th Annual Meeting of the American Economic Association, 68(2):1-16.
—. 1979. "Rational Decision Making in Business Organisations." *American Economic Review* 69(4):493-513.
—. 1990. "Invariants of Human Behavior." *Annual Review of Psychology* 41:1-19.
Simpson, Jeffry A. 1990. "Influence of Attachment Styles on Romantic Relationships." *Journal of Personality and Social Psychology* 59(5):971-980.
Singer, Tania, Stefan J. Kiebel, Joel S. Winston, Raymond J. Dolan, and Chris D. Frith. 2004. "Brain Responses to the Acquired Moral Status of Faces." *Neuron* 41:653-662.
Sitkin, Sim B., and Naney L. Roth. 1993. "Explaining the Limited Effectiveness of Legalistic 'Remedies' for Trust/Distrust." *Organization Science* 4(3):367-392.
Six, Frédérique. 2005. *The Trouble with Trust: The Dynamics of Interpersonal Trust Building*. Cheltenham, UK; Northampton, MA: Edward Elgar.
Sloman, S. A:. 1996. "The Empirical Case for Two Systems of Reasoning." *Psychological Bulletin* 119(1):3-22.
Slonim, R., and A. E. Roth. 1998. "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic." *Econometrica* 66(3):569-596.
Slovic, P. 1993. "Perceived Risk, Trust, and Democracy." *Risk Analysis* 13(6):675-682.
Slovic, P., M. Finucane, E. Peters, and D. G. MacGregor. 2002. "The Affect Heuristic." Pp. 397-420 in *Heuristics and Biases*, edited by T: Gilovich, D: Griffin, and D. Kahneman. New York: Cambridge University Press.
Slovic, Paul. 1995. "The Construction of Preference." *American Psychologist* 50(5):364-371.
Smelser, Neil J. 1992. "The Rational Choice Perspective: A Theoretical Assessment." *Rationality and Society* 4(4):381-410.
Smith, Craig A., and Phoebe C. Ellsworth. 1985. "Patterns of Cognitive Appraisal in Emotion." *Journal of Personality and Social Psychology* 48(4):813-838.
Smith, Edward E. 1968. "Choice Reaction Time: An Analysis of Major Theoretical Positions." *Psychological Bulletin* 69(2):77-110.

Smith, Eliot R., and Jamie DeCoster. 2000. "Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems." *Personality and Social Psychology Review* 4(2):108-131.
Smith, Eliot R., and Gün R. Semin. 2004. "Socially Situated Cognition: Cognition in its Social Context." *Advances in Experimental Social Psychology* 36:53-117.
Smith, Stephen M., and Irwin P. Levin. 1996. "Need for Cognition and Choice Framing Effects." *Journal of Behavioral Decision Making* 9(4):283-290.
Sonnemans, Joep, Arthur Schram, and Theo Offerman. 1998. "Public Good Provision and Public Bad Prevention: The Effect of Framing." *Journal of Economic Behavior & Organization* 34(1):143-161.
Spitzer, Manfred, Urs Fischbacher, Bärbel Herrnberger, Georg Grön, and Ernst Fehr. 2007. "The Neural Signature of Norm Compliance." *Neuron* 56:185-196.
Stanovich, K. E. 2004. *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. Chicago: Chicago University Press.
Stocké, V. 2002. *Framing und Rationalität. Die Beeutung der Informationsdarstellung für das Entscheidungsverhalten*. München: Oldenbourg.
Stocké, Volker. 2006. "Attitudes toward Surveys, Attitude Accessibility and the Effect on Respondents' Susceptibility to Nonresponse." *Quality & Quantity* 40(2):259-288.
—. 2007a. "Explaining Educational Decision and Effects of Families' Social Class Position: An Empirical Test of the Breen–Goldthorpe Model of Educational Attainment." *European Sociological Review* 23(4):505-519.
—. 2007b. "The Interdependence of Determinants for the Strength and Direction of Social Desirability Bias in Racial Attitude Surveys." *Journal of Official Statistics* 23(4):493-514.
Strack, F., and R. Deutsch. 2004. "Reflective and Impulsive Determinants of Social Behavior." *Personality and Social Psychology Review* 8(3):220-247.
Strack, Fritz, and Roland Deutsch. 2009. "Intuition." Pp. 179-198 in *Social Cognition. The Basis of Human Interaction*, edited by Fritz Strack, and Jens Förster. New York, London: Psychology Press.
Stryker, S., and A. Statham. 1985. "Symbolic Interactionism and Role Theory." Pp. 311-378 in *Handbook of Social Psychology*, edited by G. Lindzey, and E. Aronson. New York: Random House.
Stryker, Sheldon, and Peter J. Burke. 2000. "The Past, Present, and Future of Identity Theory." *Social Psychology Quarterly* 63(4):284-297.
Sztompka, Piotr. 1996. "Trust and Emerging Democracy." *International Sociology* 11(1):37-62.
—. 1998. "Trust, Distrust and Two Paradoxes of Democracy." *European Journal of Social Theory* 1(1):19-32.
—. 1999. *Trust: A Sociological Theory*. Cambridge: Cambridge University Press.
Tajfel, H. 1982. "Social Psychology of Intergroup Relations." *Annual Review of Psychology* 33:1-39.
Tajfel, H., and J. C. Turner. 1979. "An Integrative Theory of Intergroup Conflict." Pp. 33-47 in *The Social Psychology of Intergroup Relations*, edited by W.G. Austin, and P. Worchel. Monterey: Brooks/Cole.
Tan, Jonathan H.W., and Claudia Vogel. 2008. "Religion and Trust: An Experimental Study." *Journal of Economic Psychology* 29(6):832-848.
Tanis, Martin, and Tom Postmes. 2005. "Short Communication: A Social Identity Approach to Trust: Interpersonal Perception, Group Membership and Trusting Behavior." *European Journal of Social Psychology* 35(3):413-424.
Thoits, Peggy A., and Lauren K. Virshup. 1997. "Me´s and We´s. Forms and Functions of Social Identities." Pp. 106-133 in *Self and Identitiy. Fundamental Issues*, edited by R. D. Ashmore, and L. Jussim. Oxford, New York: Oxford University Press.
Thom, David H., and Bruce Campbell. 1997. "Patient-Physician Trust: An Exploratory Study." *Journal of Family Practice* 44(2):169-176.
Thomas, William Isaac, and Florian Znaniecki. 1927. *The Polish Peasant in Europe and America*. New York: Alfred A. Knopf.
Thompson, Valerie A. 2009. "Dual-Process Theories: A Meta-Cognitive Perspective." Pp. 171-195 in *In Two Minds. Dual Processes and Beyond*, edited by J. S. B. T. Evans, and K. Frankish. New York: Oxford University Press.
Todorov, Alexander, Manish Pakrashi, and Nikolas N. Oosterlof. 2009. "Evaluating Faces on Trustworthiness After Minimal Time Exposure." *Social Cognition* 27(6):813-833.
Tomlinson, Edward C., and Roger C. Mayer. 2009. "The Role of Causal Attribution Dimensions in Trust Repair." *American Sociological Review* 34(1):85-104.
Tooby, John, and Leda Cosmides. 1992. *The Psychological Foundations of Culture*. New York: Oxford University Press.
Townsend, James T., and F. Ashby. 1983. *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge, NY: Cambridge Univerity Press.

Trevino, Linda K., Robert H. Lengel, and Richard L. Daft. 1987. "Media Symbolism, Media Richness, and Media Choice in Organizations." *Communication Research* 14(5):553-574.
Triandis, Harry C. 1989. "The Self and Social Behavior in Differing Cultural Contexts." *Psychological Review* 96(3):506-520.
—. 1995. *Individualism and Collectivism*. Boulder, CO: Westview.
Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46(1):35-57.
Turner, J. C., M. Hogg, P. J. Oakes, S. Reicher, and M. Wetherell. 1987. *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Basil Blackwell.
Turner, John C., Penelope J. Oakes, Alexander S. Haslam, and Craig McGarty. 1994. "Self and Collective: Cognition and Social Context." *Personality and Social Psychology Bulletin* 20(5):454-463.
Tversky, A., and D. Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5(2):207-232.
Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Play." *Science* 211(4481):453-458.
—. 1983. "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90(4):293-315.
—. 1986. "Rational Choice and the Framing of Decisions." *Journal of Business* 59(4):S251-S278.
Tversky, Amos, and Itamar Simonson. 1993. "Context-Dependent Preferences." *Management Science* 39(10):1179-1189.
Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review* 80(1):204-217.
Tyler, Tom R. 2001. "Why do People Rely on Others? Social Identity and the Social Aspects of Trust." Pp. 285-306 in *Trust in Society*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Ulrich, R., and J. Miller. 1994. "Effects of Truncation of Reaction Time Data." *Journal of Experimental Psychology* 123(1):34-80.
Uzzi, Brian. 1997. "Social Structure and Competition in Interfirm Networks: The Paradox of Embededness." *Administrative Science Quarterly* 42(1):35-67.
Valley, Kathleen, Joseph Moag, and Max Bazerman. 1998. ""A Matter of Trust": Effects of Communication on the Efficiency and Distribution of Outcomes." *Journal of Economic Behavior & Organization* 34(2):211-238.
Van Zandt, Trisha. 2000. "How to Fit a Response Time Distribution." *Psychonomic Bulletin & Review* 7(3):424-463.
Van Zandt, Trisha, and Roger Ratcliff. 1995. "Statistical Mimicking of Reaction Time Data: Single-Process Models, Parameter Variability, and Mixtures." *Psychonomic Bulletin & Review* 2(1):20-54.
Vanberg, Christoph. 2008. "Why do People Keep Their Promises? An Experimental Test of Two Explanations." *Econometrica* 76(6):1467-1480.
Vanberg, Viktor J. 1994. *Rules and Choice in Economics*. London: Routledge.
—. 2002. "Rational Choice versus Program Based Behavior: Alternative Theoretical Approaches and their Relevance for the Study of Institutions." *Rationality and Society* 14(1):7-54.
—. 2004. "The Rationality Postulate in Economics: Its Ambiguity, its Deficiency and its Evolutionary Alternative." *Journal of Economic Methodology* 11(1):1-29.
Visser, Penny S., George Y. Bizer, and Jon A. Krosnick. 2006. "Exploring the Latent Construct of Strength-Related Attitude Atttributes." *Advances in Experimental Social Psychology* 38:1-67.
Vohs, Kathleen D., Nicole L. Mead, and Miranda R. Goode. 2006. "The Psychological Consequences of Money." *Science* 314:1154-1156.
von Neumann, John, and Oskar Morgenstern. 1944. *A Theory of Games and Economic Behavior*, 2 ed. Princeton: Princeton University Press.
Watzlawick, P., J. H. Beavin, and D. D. Jackson. 1967. *Pragmatics of Human Communication*. New York: W. W. Norton.
Weber, J. Mark, Shirli Kopelman, and David M. Messick. 2004. "A Conceptual Review of Decision Making in Social Dilemmas: Applying a Logic of Appropriateness." *Personality and Social Psychology Review* 8(3):281-307.
Weber, Linda R., and Allison I. Carter. 2003. *The Social Construction of Trust*. New York: Kluwer/Plenum Publishers.
Weesie, Jeroen, and Werner Raub. 1996. "Private Ordering: A Comparative Institutional Analysis of Hostage Games." *Journal of Mathematical Sociology* 21(3):201-240.
Whalen, Paul J., Scott L. Rauch, Nancy L. Etcoff, Sean C. McInerney, Michael B. Lee, and Michael A. Jenike. 1998. "Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity Without Explicit Knowledge." *Journal of Neuroscience* 18(1):411-418.

Wheeler, Christian S., and Richard E. Petty. 2001. "The Effects of Stereotype Activation on Behavior: A Review of Possible Mechanisms." *Psychological Bulletin* 127(6):797-826.
Whitener, Ellen M., Susan E. Brodt, Audrey M. Korsgaard, and Jon M. Werner. 1998. "Managers As Initiators of Trust: An Exchange Relationship Framework for Understanding Managerial Trustworhy Behavior." *Academy of Management Journal* 23(3):513-530.
Williams, Michelle. 2001. "In Whom We Trust: Group Membership as an Affective Context for Trust Development." *The Academy of Management Review* 26(3):377-396.
—. 2007. "Building Genuine Trust through Interpersonal Emotion Management: A Threat Regulation Model of Trust and Collaboration Across Boundaries." *The Academy of Management Review* 32(2):595-621.
Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: The Free Press.
—. 1993. "Calculativeness, Trust, and Economic Organization." *The Journal of Law and Economics* 36(1):453-486.
Wilson, J. M., S. G. Straus, and B. Mcevily. 2006. "All in Due Time: The Development of Trust in Computer-Mediated and Face-to-Face Teams." *Organizational Behavior and Human Decision Processes* 99(1):16-33.
Wilson, Rick K., and Catherine C. Eckel. 2006. "Judging a Book by its Cover: Beauty and Expectations in the Trust Game." *Political Research Quarterly* 59(2):189-202.
Winkielman, P., N. Schwarz, T. Fazendeiro, and R. Reber. 2003. "The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment." Pp. 189-217 in *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, edited by J. Musch, and K. C. Klauer. Mahwah, NJ: Lawrence Erlbaum Associates.
Winkielman, Piotr, and Jonathan W. Schooler. 2009. "Uncinscious, Conscious, and Metaconscious in Social Cognition." Pp. 49-70 in *Social Congition. The Basis of Human Interaction*, edited by Fritz Strack, and Jens Förster. New York, London: Psychology Press.
Winston, J.S., B.A. Strange, J. O´Doherty, and R.J. Dolan. 2002. "Automatic and Intentional Brain Responses During Evaluation of Trustworthiness of Faces." *Nature Neuroscience* 5(3):277-283.
Woolcock, Michael. 1998. "Social Capital and Economic Development: Toward a Theoretical Synthesis and Policy Framework." *Theory and Society* 27(2):151-208.
Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.
Worchel, P. 1979. "Trust and Distrust." Pp. 174-187 in *The Social Psychology of Intergroup Relations*, edited by W.G. Austin, and S. Worchel. Belmont, CA: Wadsworth.
Wright, William F. , and Urton Anderson. 1989. "Effects of Situation Familiarity and Financial Incentives on Use of the Anchoring and Adjustment Heuristic for Probability Assessment." *Organizational Behavior and Human Decision Processes* 44(1):68-82.
Wrightsman, L. S. 1974. *Assumptions About Human Nature. A Social-Psychological Approach*. Monterey, CA: Brooks/Cole.
—. 1991. "Interpersonal Trust and Attitudes Toward Human Nature." Pp. 373-412 in *Measures of Personality and Social Psychological Attitudes. Volume 1: Measures of Social Psychological Attitudes*, edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightsman. San Diego: Academic Press.
Yamagishi, Toshio. 2001. "Trust as a Form of Social Intelligence." Pp. 121-147 in *Trust in Society.*, edited by Karen S. Cook. New York: Russel Sage Foundation.
Yamagishi, Toshio, Satoshi Kanazawa, Rie Mashima, and Shigeru Terai. 2005. "Separating Trust from Cooperation in a Dynamic Relationship: Prisoner´s Dilemma with Variable Dependence." *Rationality and Society* 17(3):275-308.
Yamagishi, Toshio, Masako Kikuchi, and Motoko Kosugi. 1999. "Trust, Gullibilty, and Social Intelligence." *Asian Journal of Social Psychology* 2:145-161.
Yamagishi, Toshio, Shigeru Terai, Toko Kiyonari, Nobuhiro Mifune, and Satoshi Kanazawa. 2007. "The Social Exchange Heuristic: Managing Errors in Social Exchange." *Rationality and Society* 19(3):259-291.
Yamagishi, Toshio, and Midori Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18(2):129-166.
Zaheer, Akbar, and Bill McEvily. 1998. "Does Trust Matter? Exploring the Effects of Interorganizational and Interpersonal Trust on Performance." *Organization Science* 9(2):141-159.
Zak, Paul J. 2004. "Neuroeconomics." *Philosophical Transactions of the Royal Society London* 359:1737-1748.
—. 2005. "The Neuroeconomics of Trust." in *Working Paper Series*. Center for Neuroeconomics Studies, Claremont Graduate University.
—. 2007. "The Neuroeconomics of Trust." Pp. 17-33 in *Renaissance in Behavioral Economics*, edited by Roger Frantz. New York: Routledge.
Zak, Paul J., and Stephen Knack. 2001. "Trust and Growth." *The Economic Journal* 111(470):295-321.

Zak, Paul J., and Jager Kugler. 2011. "Neuroeconomics and International Studies: A New Understanding of Trust." *International Studies Perspectives* 12(3):136-152.
Zand, Dale E. 1972. "Trust and Managerial Problem Solving." *Administrative Science Quarterly* 17(2):229-239.
Ziegler, Rolf. 1998. "Trust and the Reliability of Expectations." *Rationality and Society* 10(4):427-450.
Zucker, Lynne G. 1986. "Production of Trust: Institutional Sources of Economic Structure: 1840-1920." Pp. 53-111 in *Research in Organizational Behavior, Vol. 8*, edited by Barry M. Staw, and Larry L. Cummings. Greenwich, London: JAI Press.

Appendix A: Omitted Tables and Results

Chapter 6.5.2, Table 10: Trust and chronic script accessibility, omitted control variables

	Tobit	Robust	GLM
Nfcscale	0.181	0.168	0.744
	(1.02)	(1.14)	(1.2)
Fiscale	-0.0577	-0.0857	-0.217
	(-0.29)	(-0.52)	(-0.33)
Append	-0.0775+	-0.0711+	-0.27
	(-1.47)	(-1.52)	(-1.44)
age3	-0.0142+	-0.0117+	-0.0435
	(-1.48)	(-1.57)	(-1.29)
Sex	-0.0736+	-0.0719+	-0.311*
	(-1.44)	(-1.52)	(-1.71)
Partner	-0.00701	0.0164	0.0338
	(-0.15)	(0.41)	(0.21)

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Chapter 6.5.2, Table 10: Trust and chronic script accessibility, orthogonal models

	Tobit	Robust	GLM
end	-0.117***	-0.118***	-0.473***
	(-2.60)	(-2.99)	(-2.98)
frame	0.002	-0.007	-0.017
	(0.05)	(-0.17)	(-0.10)
recskala	0.204	0.203	0.805
	(0.7)	(0.8)	(0.76)
*end*recscale*	1.253**	1.001**	4.427**
	(2.15)	(2.03)	(2.07)
*frame*recscale*	0.0421	-0.0347	-0.0394
	(0.06)	(-0.06)	(-0.02)
*end*frame*	1.179*	0.857+	3.841*
	(1.84)	(1.53)	(1.65)
*end*frame*recscale*	-1.820*	-1.336+	-5.960*
	(-1.93)	(-1.63)	(-1.76)
trustscale	0.310+	0.281*	1.118+
	(1.46)	(1.67)	(1.63)
nfcscale	0.181	0.168	0.744
	(1.02)	(1.14)	(1.2)
fiscale	-0.058	-0.086	-0.217
	(-0.29)	(-0.52)	(-0.33)
append	-0.078+	-0.071+	-0.27
	(-1.47)	(-1.52)	(-1.44)
age	-0.014+	-0.012+	-0.044
	(-1.48)	(-1.57)	(-1.29)
sex	-0.0736+	-0.0719+	-0.311*
	(-1.44)	(-1.52)	(-1.71)
partner	-0.007	0.016	0.034

	(-0.15)	(0.41)	(0.21)
constant	0.166	0.189	-1.333
	(0.56)	(0.76)	(-1.26)
Pseudo R^2 (ps. *LL*)	0.084	0.1105	(-155.2)
Wald (full model)	36.28***	44.05***	33.76***
χ^2 Improvement (4*df*)	7.9*	7.38+	7.23+

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Chapter 6.5.2, Table 11: Trust and chronic frame accessibility, omitted control variables

	Tobit	Robust	GLM
fiscale	-0.076	-0.088	-0.26
	(-0.37)	(-0.50)	(-0.38)
nfcscale	0.202	0.21	0.864
	(1.14)	(1.42)	(1.4)
append	-0.059	-0.058	-0.218
	(-1.07)	(-1.20)	(-1.13)
age	-0.015+	-0.014*	-0.047
	(-1.52)	(-1.73)	(-1.33)
sex	-0.076	-0.072+	-0.308*
	(-1.39)	(-1.50)	(-1.67)
partner	-0.006	0.019	0.034
	(-0.13)	(0.48)	(0.22)

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Chapter 6.5.2, Table 11: Trust and chronic frame accessibility, orthogonal models

	Tobit	Robust	GLM
end	-0.117**	-0.120***	-0.463***
	(-2.53)	(-2.96)	(-2.89)
frame	0.004	0.0003	-0.01
	(0.09)	(0.01)	(-0.06)
trustscale	0.249	0.256	0.926
	(0.98)	(1.18)	(1.15)
*end*frame*	-0.339	-0.123	-1.067
	(-0.66)	(-0.28)	(-0.64)
*end*trustscale*	-0.301	0.001	-0.392
	(-0.54)	(0)	(-0.21)
*frame*trustscale*	-0.231	0.001	-0.517
	(-0.34)	(0)	(-0.24)
*end*frame*trustscale*	0.483	0.122	1.444
	(0.56)	(0.17)	(0.52)
recscale	0.296	0.292	1.066
	(1.01)	(1.15)	(1.04)
fiscale	-0.076	-0.088	-0.26
	(-0.37)	(-0.50)	(-0.38)
nfcscale	0.202	0.21	0.864
	(1.14)	(1.42)	(1.4)

append	-0.059	-0.058	-0.218
	(-1.07)	(-1.20)	(-1.13)
age	-0.015+	-0.014*	-0.047
	(-1.52)	(-1.73)	(-1.33)
sex	-0.076	-0.072+	-0.308*
	(-1.39)	(-1.50)	(-1.67)
partner	-0.006	0.019	0.034
	(-0.13)	(0.48)	(0.22)
constant	0.123	0.097	-1.506
	(0.39)	(0.37)	(-1.39)
Pseudo R^2 (Ps. *LL*)	0.0652	0.0945	-156.54
Wald (full model)	22.31*	29.15***	22.4**
χ^2 Improvement (4*df*)	0.75	0.55	0.81

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Chapter 6.5.2, Table 12: Trust and activation weight components, omitted control variables

	Tobit	Robust	GLM
nfcscale	0.171	0.17	0.743
	-0.94	-1.12	-1.16
fiscale	-0.0585	-0.0765	-0.218
	(-0.29)	(-0.46)	(-0.33)
append	-0.0718	-0.0667	-0.251
	(-1.35)	(-1.40)	(-1.32)
age	-0.0140+	-0.0116+	-0.0426
	(-1.45)	(-1.54)	(-1.26)
sex	-0.0842*	-0.0813*	-0.342*
	(-1.65)	(-1.70)	(-1.87)
partner	-0.00132	0.0209	0.0455
	(-0.03)	-0.52	-0.28

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Chapter 6.5.2: Table 12: Trust and activation weight components, orthogonal models

	Tobit	Robust	GLM
frame	0.003	-0.007	-0.012
	(0.06)	(-0.17)	(-0.07)
trustscale	0.330+	0.308*	1.198*
	(-1.52)	(1.84)	(1.69)
recscale	0.304	0.27	1.069
	(1.06)	(1.09)	(1.03)
*frame*trustscale*	-0.153	-0.06	-0.359
	(-0.39)	(-0.19)	(-0.27)
*frame*recscale*	-0.917*	-0.736*	-3.025*
	(-1.83)	(-1.69)	(-1.69)
*trustscale*recscale*	2.933	1.818	6.193
	(0.95)	(0.75)	(0.6)
*frame*trust.*rec**	-1.362	0.213	0.237
	(-0.28)	(0.05)	(0.01)

end	-0.116**	-0.118***	-0.460***
	(-2.51)	(-2.91)	(-2.85)
nfcscale	0.181	0.175	0.77
	(0.99)	(1.13)	(1.2)
fiscale	-0.054	-0.075	-0.205
	(-0.27)	(-0.44)	(-0.31)
append	-0.070	-0.066	-0.245
	(-1.32)	(-1.40)	(-1.28)
age	-0.014+	-0.012+	-0.044
	(-1.49)	(-1.58)	(-1.32)
sex	-0.082+	-0.079*	-0.335*
	(-1.60)	(-1.66)	(-1.82)
partner	-0.008	0.016	0.024
	(-0.17)	(0.39)	(0.15)
constant	0.0835	0.115	-1.591+
	(0.28)	(0.46)	(-1.54)
Ps. R^2 (ps. *LL*)	0.076	0.105	(-155.9)
Wald (full model)	31.34***	36.68***	28.05**
χ^2 Improvement (4*df*)	4.19	3.56	3.18

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

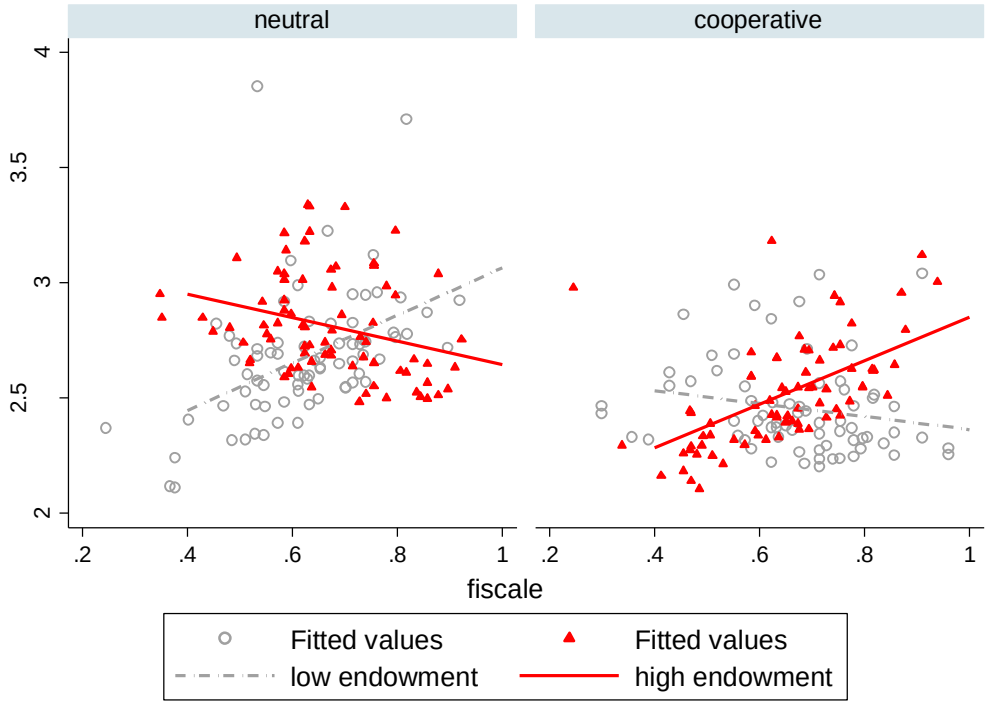
Chapter 6.5.3: Regression of *reltrust* on processing preferences

Variable	Orthogonal, using *nfcscale*			*Variable*	Orthogonal, using *fiscale*		
	Tobit	Robust	GLM[1)]		Tobit	Robust	GLM[1)]
end	-0.121***	-0.124***	-0.475***	*end*	-0.117**	-0.123***	-0.469***
	(-2.64)	(-3.10)	(-3.01)		(-2.45)	(-3.03)	(-2.90)
frame	0.003	-0.002	-0.021	*frame*	0.004	-0.003	-0.014
	(-0.05)	(-0.06)	(-0.13)		(-0.08)	(-0.07)	(-0.08)
nfcscale	0.122	0.152	0.62	*fiscale*	-0.005	-0.029	-0.093
	-0.61	-0.88	-0.9		(-0.02)	(-0.14)	(-0.12)
*end*frame*	-0.886+	-0.51	-2.75	*end*frame*	0.359*	0.321*	1.268*
	(-1.51)	(-1.03)	(-1.34)		(-1.69)	(-1.87)	(-1.86)
*end*nfcscale*	-0.385	-0.192	-1.189	*end*fiscale*	0.312	0.283	1.086
	(-0.87)	(-0.49)	(-0.73)		(-1.12)	(-1.2)	(-1.13)
*frame*nfcscale*	-0.697	-0.409	-1.926	*frame*fi.*	0.253	0.244	0.741
	(-1.34)	(-0.92)	(-1.09)		(-0.47)	(-0.55)	(-0.42)
*end*frame.*	1.073+	0.601	3.279	*end*frame*	-0.192	-0.0253	-0.434
** nfcscale*	(-1.45)	(-0.97)	(-1.28)	** fiscale*	(-0.37)	(-0.06)	(-0.26)
trustscale	0.268	0.237	0.975+	*trustscale*	0.225	0.135	0.622
	(-1.3)	(-1.41)	(-1.45)		(-0.42)	(-0.3)	(-0.36)
recscale	0.257	0.248	0.938	*recscale*	-0.474	-0.448	-1.475
	(-0.85)	(-0.97)	(-0.91)		(-0.60)	(-0.68)	(-0.56)
age	-0.066	-0.057	-0.234	*age*	-0.068	-0.064	-0.246
	(-1.23)	(-1.21)	(-1.24)		(-1.28)	(-1.41)	(-1.32)
sex	-0.014+	-0.013*	-0.045	*sex*	-0.016*	-0.014*	-0.05+
	(-1.46)	(-1.75)	(-1.31)		(-1.70)	(-1.94)	(-1.47)
partner	-0.082+	-0.079+	-0.342*	*partner*	-0.089*	-0.085*	-0.361**
	(-1.48)	(-1.60)	(-1.84)		(-1.67)	(-1.86)	(-2.00)
append	-0.016	0.013	0.007	*append*	-0.004	0.018	0.042
	(-0.34)	(-0.33)	(-0.04)		(-0.09)	(-0.46)	(-0.27)
constant	0.167	0.136	-1.373	*constant*	0.17	0.208	-1.106
	(-0.53)	(-0.51)	(-1.28)		(-0.65)	(-0.96)	(-1.26)

Ps. R^2 (ps. *LL*)	0.07	0.959	(-156.3)	0.067	0.095	(-156.68)
Wald (full model)	25.56**	31.17***	25.85**	22.5**	29.7***	22.14*
χ^2 Improvement (4*df*)	2.52	1.24	2.03	1.94	1.69	1.54

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. [1)] Effects on log-odds. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Chapter 6.6.4: Omitted figure displaying predicted *logtime*, using model specification (4)



Chapter 6.7.3: Combining accessibility and processing preferences, model specification (6)

Variable	Tobit		Robust		GLM[1)]	
end	-0.100**	-0.104**	-0.110**	-0.111***	-0.409**	-0.435***
	(-2.07)	(-2.20)	(-2.57)	(-2.71)	(-2.52)	(-2.65)
frame	-0.00165	0.0157	-0.0123	-0.00306	-0.0116	0.0369
	(-0.03)	-0.33	(-0.30)	(-0.07)	(-0.07)	-0.22
recscale	0.567*	0.880***	0.384+	0.667**	1.955*	3.049**
	-1.79	-2.58	-1.46	-2.38	-1.74	-2.44
fiscale	0.011	-0.0537	-0.0284	-0.0863	0.0235	-0.14
	-0.05	(-0.23)	(-0.13)	(-0.44)	-0.03	(-0.18)
*end*frame*	10.28***	10.90***	7.018**	7.129**	33.01**	36.58***
	-2.77	-2.9	-2.45	-2.42	-2.56	-2.62
*end*fiscale*	8.641**	9.288***	5.224**	6.129**	25.24**	28.25**
	-2.53	-2.62	-2.17	-2.34	-2.11	-2.21
*frame*fiscale*	2.076	2.593	2.207	2.909	7.376	9.449
	-0.61	-0.73	-0.71	-0.86	-0.65	-0.75
*end*recscale*	9.751***	10.19***	6.154***	6.829***	28.94**	31.22**
	-2.89	-2.94	-2.65	-2.75	-2.49	-2.52
*frame*recscale*	1.926	2.104	2.047	2.429	6.962	7.699
	-0.58	-0.63	-0.69	-0.79	-0.64	-0.67
*recscale*fiscale*	1.81	3.051	1.337	2.564	5.118	10.06
	-0.61	-0.94	-0.53	-0.9	-0.51	-0.87
end frame*recscale*	-15.58***	-16.04***	-10.82**	-10.59**	-49.91**	-53.59**
	(-2.78)	(-2.84)	(-2.49)	(-2.39)	(-2.57)	(-2.56)
*end*frame*fiscale*	-14.55***	-15.60***	-9.832**	-10.30**	-46.48**	-52.29**
	(-2.59)	(-2.72)	(-2.23)	(-2.22)	(-2.38)	(-2.45)
*end*recscale*fiscale*	-13.51***	-14.30***	-8.253**	-9.404**	-39.13**	-43.01**
	(-2.60)	(-2.69)	(-2.24)	(-2.39)	(-2.18)	(-2.26)
*frame*rec.*fis.*	-3.117	-3.472	-3.395	-4.136	-11.39	-12.87
	(-0.61)	(-0.65)	(-0.72)	(-0.82)	(-0.66)	(-0.69)
*end*frame*rec.*fis.*	21.85***	22.76***	15.00**	15.13**	69.67**	75.91**
	(2.58)	-2.64	-2.25	-2.18	-2.37	-2.38
constant	0.0932	0.0542	0.0542	0.185	-1.414+	-1.806
	-0.34	-0.14	-0.14	-0.61	(-1.48)	(-1.36)
Pseudo R^2 (ps. LL)	0.076	0.126	0.084	0.141	(-156.58)	(-152.81)
Wald (full model)	22.83**	42.84***	28.85**	47.07***	19.4	35.64**
χ^2 Improvement (11*df*)	15.31+	15.07	16.56+	14.09	12.63	12.08
Control variables	No	Yes	No	Yes	No	Yes

Note: N=298 observations in all models. T-values in brackets. All models use non-parametric bootstrapping of parameter estimates with 2000 replications. [1)] Effects on log-odds. + p<0.15, * p<0.10, ** p<0.05, *** p<0.01.

Ad table 16: Fitting DT distributions across subgroups

	Fitting Across Experimental Conditions (outliers excluded, N=289)							
	Low/Neutral		Low/Cooperative		High/Neutral		High/Cooperative	
	D=	p=	D=	p=	D=	p=	D=	p=
Lognormal	0.084	0.638	0.088	0.638	0.081	0.710	0.059	0.945
Log-Logistic	0.083	0.646	0.068	0.892	0.078	0.744	0.078	0.715
Inv. Gauss	0.074	0.791	0.081	0.734	0.085	0.641	0.059	0.944
Weibull	0.113	0.283	0.135	0.149+	0.110	0.322	0.081	0.683

Appendix B: Items, Scales, and Instructions

The following tables list the items of those scales which were used in the experiment. They also present all associated measures of reliability. All scales were elicited using a 7-point Likert-type scale ranging from “fully agree” to “fully disagree,” including a “don´t know”-option. The reliability measures obtained refer to the full data sample including N=298 observations. The scales were constructed by computing the average row mean across all items of the scale, normalizing the scale range to [0,1]. Missing values were left out. A list of items translated into English is available from the author on request.

1. Interpersonal Trust Inventory (Kassebaum 2004), short version

Item	Factor Loading
(1) In der Regel begegne ich fremden Menschen mit großer Vorsicht	0.2564
(2) Die meisten Menschen würden eine günstige Gelegenheit nutzen, um sich auf Kosten anderer zu bereichern	0.6561
(3) Ich gehe in der Regel davon aus, dass andere Menschen mir gegenüber nicht nur gute Absichten haben	0.6724
(4) Institutionen wie Verwaltungen, Behörden, Ämtern usw., kann ich nur sehr schwer vertrauen	0.5265
(5) Ich habe oft Angst davor, dass fremde Menschen mir und meiner Umwelt Schaden zufügen könnten.	0.5159
(6) Im Grunde kann man den Mitmenschen vertrauen.	0.4123
(7) Wenn man seine finanziellen Angelegenheiten nicht weitgehend selbst regelt, muss man befürchten, hereingelegt oder hintergangen zu werden.	0.4511
(8) Manchmal befürchte ich, dass sogenannte "Experten" Entscheidungen treffen könnten, die sich negativ auf mein Wohlergehen auswirken	0.5126
(9) Wenn andere eine Aufgabe für mich erledigen, würde ich mich am liebsten ständig vergewissern, ob sie es auch in meinem Sinne und nach meinen Vorstellungen tun.	0.5078
Factors retained: 1 Eigenvalue= 2.32 Cronbach´s Alpha: 0.77	

2. Reciprocity Scale (Perugini et al. 2003)

Item	Factor1 (pos. rec.)	Factor 2 (neg. rec.)
(1) Jemandem zu helfen ist die beste Methode um sicherzustellen, dass man in Zukunft auch selbst Hilfe erhält.	0.4663	0.1154

(2) Wenn mir jemand einen Gefallen tut, bin ich bereit, dies zu erwidern.	0.575	0.2361
(3) Wenn mir schweres Unrecht zuteil wird, werde ich mich um jeden Preis bei der nächsten Gelegenheit rächen.	-0.4957	0.6467
(4) Wenn mich jemand in eine schwierige Lage bringt, werde ich das Gleiche mit ihm machen.	-0.5448	0.617
(5) Ich strenge mich besonders an, um jemandem zu helfen, der mir früher schon geholfen hat.	0.5888	0.3426
(6) Wenn ich jemandem ein Kompliment mache, erwarte ich auch, dass er es erwidert.	-0.1011	0.3017
(7) Ich bin bereit, Kosten auf mich zu nehmen, um jemandem zu helfen, der mir früher schon einmal geholfen hat.	0.5258	0.3354
(8) Ich vermeide es, unhöflich zu sein, weil ich nicht will, dass andere unhöflich zu mir sind.	0.4002	0.0923
(9) Wenn mich jemand beleidigt, werde ich mich ihm gegenüber auch beleidigend verhalten.	-0.2973	0.369
(10) Wenn ich hart arbeite, erwarte ich einen entsprechenden Lohn.	0.2116	0.3542
(11) Wenn mich jemand höflich nach etwas fragt, helfe ich gerne weiter.	0.6541	0.0867
(12) Wenn mir jemand die richtigen Lottozahlen nennt, gebe ich ihm sicherlich einen Teil des Gewinns.	0.3582	0.1695

Factors retained: 2
Eigenvalues: 2.573, 1.495
Cronbach´s Alpha: 0.6520

3. Faith In intuition Scale (Keller et al. 2000)

Item	Factor Loading
(1) Bei den meisten Entscheidungen ist es sinnvoll, sich auf sein Gefühl zu verlassen.	0.6987
(2) Ich bin ein sehr intuitiver Mensch.	0.6248
(3) Wenn es um Menschen geht, kann ich meinem unmittelbaren Gefühl vertrauen.	0.7611
(4) Ich vertraue meinen unmittelbaren Reaktionen auf andere	0.7187
(5) Der erste Einfall ist oft der beste.	0.4887
(6) Wenn die Frage ist, ob ich anderen vertrauen soll, entscheide ich normalerweise aus dem Bauch heraus.	0.589
(7) Mein erster Eindruck von anderen ist fast immer zutreffend.	0.5409
(8) Ich spüre meistens sofort, wenn jemand lügt	0.3725
(9) Wenn ich mir eine Meinung zu einer Sache bilden soll, verlasse ich mich ganz auf meine Intuition	0.5806
(10) Ich glaube, ich kann meinen Gefühlen vertrauen.	0.7095

(11) Ich kann mir über andere sehr schnell einen Eindruck bilden.	0.4972

Factors retained: 1
Eigenvalue: 4.079
Cronbach´s Alpha: 0.8462

4. Need for Cognition Scale (Keller et al. 2000)

Item	Factor Loading
(1) Ich finde es nicht sonderlich aufregend, neue Denkweisen zu erlernen.	0.4915
(2) Ich finde wenig Befriedigung darin, angestrengt stundenlang nachzudenken	0.6279
(3) Abstrakt zu denken reizt mich nicht.	0.6438
(4) Die Vorstellung, mich auf mein Denkvermögen zu verlassen, um es zu etwas zu bringen, spricht mich nicht an.	0.478
(5) Ich würde lieber etwas tun, das wenig Denken erfordert, als etwas, das mit Sicherheit meine Denkfähigkeit herausfordert.	0.7355
(6) Denken entspricht nicht dem, was ich unter Spaß verstehe.	0.6099
(7) Ich trage nicht gern die Verantwortung für eine Situation, die sehr viel Denken erfordert.	0.6588
(8) Ich versuche, Situationen vorauszuahnen und zu vermeiden, in denen die Wahr-scheinliohkeit groß ist, \par dass ich intensiv über etwas nachdenken muss.	0.5581
(9) Es genügt, dass etwas funktioniert, mit ist egal, wie oder warum.	0.5359
(10) Ich akzeptiere die Dinge meist lieber so wie sie sind, anstatt sie zu hinterfragen.	0.5924
(11) Es genügt mir, einfach die Antwort zu kennen, ohne die Gründe für die Antwort auf ein Problem zu verstehen.	0.377
(12) Wenn ich eine Aufgabe erledigt habe, die viel geistige Anstrengung erfordert hat, fühle ich mich eher erleichtert als befriedigt.	0.5053
(13) Das Denken in neuen und unbekannten Situationen fällt mir schwer.	0.6109

Factors retained: 1
Eigenvalue: 4.345
Cronbach´s Alpha: 0.8588

5. Item/Scale Intercorrelations

	recscale	*trustscale*	*fiscale*	*nfcscale*
recscale	1			
trustscale	-0.2064	1		
fiscale	0.0842	0.1672	1	
nfcscale	-0.1054	0.201	-0.0078	1

6. Instructions used in the experiment

The following instructions were used in the experiment. They are listed here in the order in which they were presented to the participants. Any reference to the two experimental manipulations, that is, the framing or incentive treatments, will be highlighted, and the alternative formulation be presented in brackets, whenever possible. This section includes the (1) general instructions which participants found in their booth, and screenshots of the actual experiment, presenting (2) welcome screen, (3) on-screen instructions of the investment game (4) the control question stage, and (5) the decision stage. An English translation of the instructions is available from the author on request.

(1) General Instructions, presented on paper when seating participants in computer booth:

Allgemeine Erklärungen für die Teilnehmer:

Auszahlungen

Sie nehmen nun an einem Experiment der Universität Mannheim teil. Im Laufe des Experimentes werden Sie Entscheidungen treffen und können dabei Punkte verdienen. Die Höhe des Betrages hängt von Ihren eigenen Entscheidungen und von den Entscheidungen anderer Teilnehmer ab.

Am Ende des Experiments wird *eine* der Aufgaben, die Sie bearbeitet haben, zufällig ausgewählt. Die Entscheidungen in dieser Aufgabe werden dann zur Berechnung der endgültigen Auszahlung herangezogen. Dazu werden die Punkte im Verhältnis 1:1 in Euro umgerechnet. Der Betrag wird am Ende der Sitzung in bar ausgezahlt.

Hinweis

Während des Experiments ist es **nicht gestattet**, mit den anderen Teilnehmern des Experiments **zu kommunizieren!** Falls Sie Fragen haben, heben Sie bitte Ihre Hand. Wir kommen dann zu Ihnen und beantworten Ihre Frage. Eine Missachtung kann zum Ausschluss führen.

Dateneingabe

Dezimalzahlen werden bei der Eingabe von Daten mit einem Punkt getrennt (z.B. 6.5).

Ablauf

Zu Beginn des Experiments werden alle Personen zufällig aufgeteilt. Dabei bilden Sie und ein Partner [ein anderer Teilnerhmer] ein Team [eine Gruppe] aus zwei Personen. Weder vor noch nach dem Experiment erfahren Sie, mit wem Sie in einem Team [einer Gruppe] waren. Ebenso wird Ihr Partner [der andere Teilnehmer] Ihre Identität nicht erfahren, d.h. alle Entscheidungen bleiben anonym.

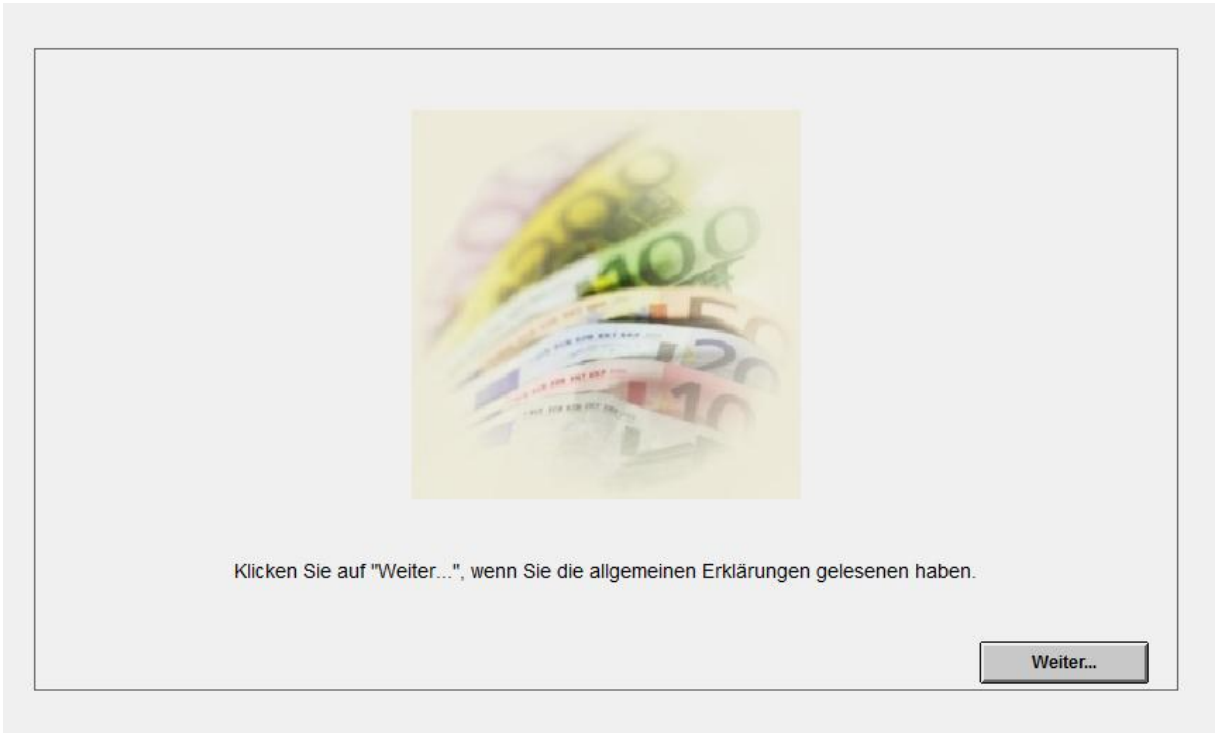
Instruktionen am Bildschirm erläutern die Aufgaben. In jeder Aufgabe treffen Sie nur eine Entscheidung. Bevor Sie eine Entscheidung treffen, können Sie deswegen die Dateneingabe üben und beantworten Kontrollfragen, die Ihnen helfen, die Aufgabe zu verstehen.

Wenn Sie diese allgemeinen Erklärungen gelesen haben, klicken Sie „Weiter…“, um mit der Bearbeitung der Aufgaben am Bildschirm zu beginnen!

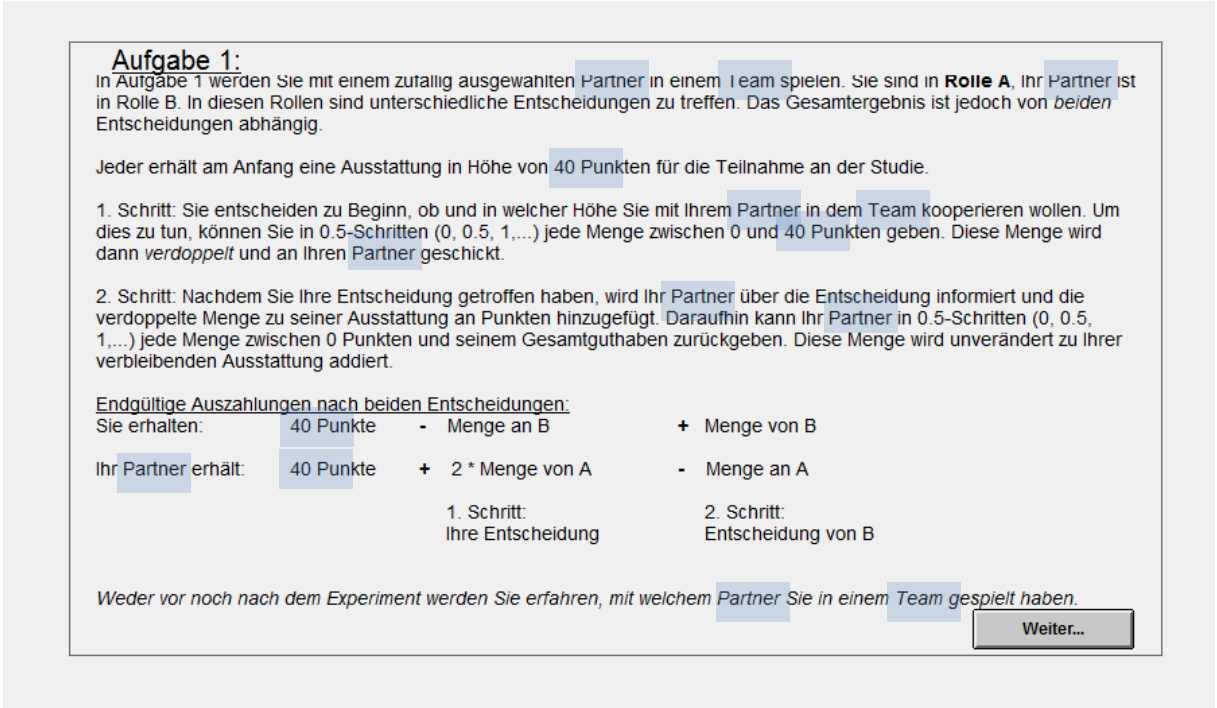
(2) Screenshot: welcome screens (presented while reading the general instructions)

(a) Cooperative framing intro screen

(b) Neutral framing intro screen

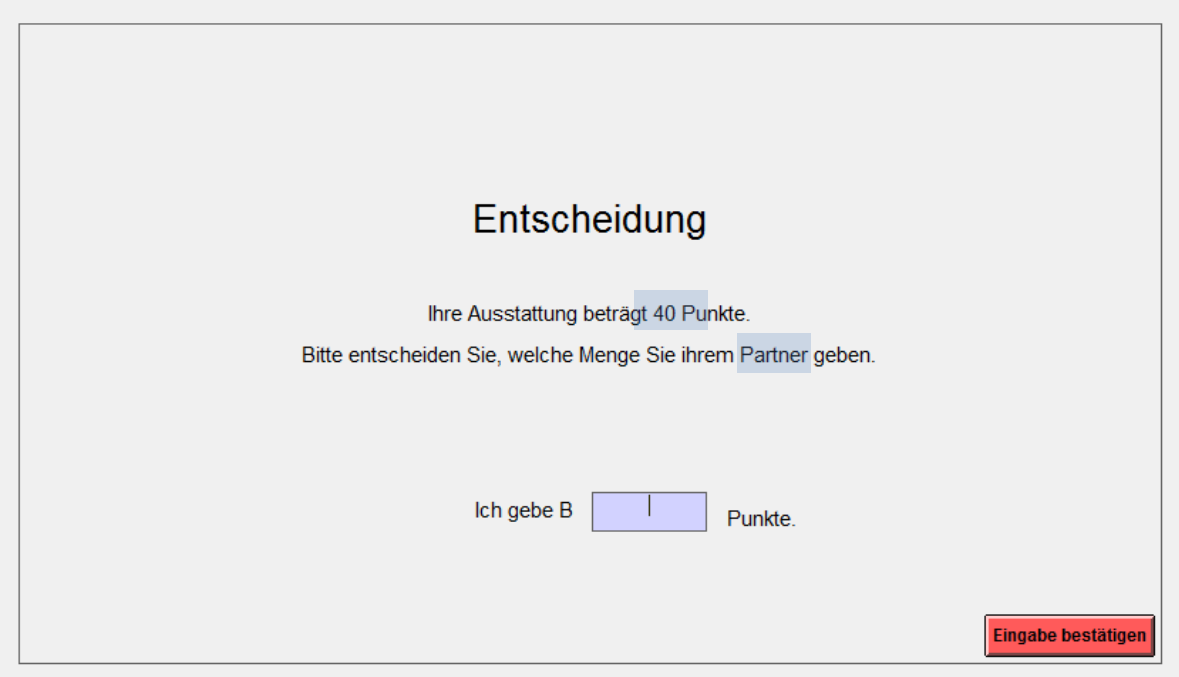


(3) Screenshot: Instructions of the investment game (cooperative framing manipulation and high incentive treatments highlighted, the neutral framing / low incentive conditions were established by replacing the fields with the corresponding formulations (i.e., "Teilnehmer" [participant], "Gruppe" [group] and low initial endowments of 7)



(4) Screenshot: On screen control questions stage (cooperative framing treatment)

(5) The decision stage of the experiment (high incentive and cooperative framing treatments highlighted and presented here)



Appendix C: Deriving Interaction Patterns

This section will demonstrate how the set of admissible *interaction patterns* that were used to guide the empirical analysis in chapter 6 can be analytically derived. The procedure can be adapted to other contexts and situations as well, by following the two steps listed below.

1. Set up bridge hypotheses

The analysis begins by linking processing modes to observable outcomes. This is the first and most important step in the analysis. Ideally, the outcome variable crucially differs between the rational and automatic processing modes. Thus, using the link, we can infer the processing mode from the observed data. In the present case, the following bridge hypotheses were used:

B1 (automatic mode): Unconditional trust leads to a complete transfer of resources, X=E.

B2 (rational mode): Conditional trust supports any transfer between zero and the initial endowment, $X_{\varepsilon}[0, E]$.

B3 (rational mode): Distrust leads to a transfer of zero, X=0.

B4 (decision time): The decision time in the automatic mode is shorter than the decision time in the rational mode.

B5 (corollary): Unconditional trust results in a shorter decision time than conditional trust.

Thus, the model predicts relatively lower transfers and relatively longer DT in the case of the rational mode, and relatively higher transfer decisions and shorter DT in the automatic mode. Of course, these bridge hypotheses can be criticized on empirical and theoretical grounds. In the present case, one can claim that rational mode decisions may lead to full trust, and likewise, that automatic mode decisions can lead trustors into distrust as well. The argument that was advanced in chapter 6.1 and 6.3 is that, *on average*, the proposed relations will hold. This proposition is based on a review of previous studies and empirical findings. Overall, this step accomplishes that the results of mode-selection (automatic mode, rational mode) are linked to the two dependent variables which are collected in the experiment.

Next, to simplify the mode-selection threshold and reduce the number of variables which are varied along with the treatment conditions, all remaining parameters should be held constant or controlled for. A number of additional bridge hypotheses have to be set up for those variables which cannot be empirically controlled. In particular, the following additional assumptions are made when testing the model:

1. A trust frame is linked to an appropriate script, that is, $a_{j|k}=1$ (A1)
2. The script regulates action to a high degree, such that $a_{k|j}=1$ (A2) (see chapters 4.6 and 4.7)
3. Situational cues are *significant* symbols with respect to indicating the appropriateness of the trust-related frame F_t, such that $l_i=1$
4. There is a potential gain involved in preventing inference errors which outweighs the costs of processing, such that $C<p*U$

An important measure to guarantee that these assumptions are valid and can be defended is the *randomization procedure* as part of any experimental design. Randomizing subjects into treatment conditions ensures that any unobserved heterogeneity is evenly distributed among all treatments and that a systematic influence can be ruled out. The statistical control of those remaining parameters for which a control measure exists adds additional information to the statistical analysis but is, in principle, not necessary.

2. Join experimental conditions and processing modes

In a second step, it is necessary to determine the potential outcome of mode-selection in each experimental condition, varying all variables under scrutiny at their potential levels. In the present case, the experimental factors vary on two levels, yielding a 2x2 between-subject design. The experimental treatments change two parameters of the threshold. First, the cooperative *versus* neutral context is designed to influence the presence of situational cues $o_{i,}$ as part of the match $m_i = m_i(o_i)$. Second, the high *versus* low incentive treatment is designed to manipulate cognitive motivation $U = (U_{rc}+C_w)$. The third parameter depends on the concrete model specification. For example, in model specification (1), the chronic accessibility of a trust-related script is varied along with the experimental treatments. What does the model tell us about the interaction between the two parameters, the interaction between each parameter and the chronic accessibility a_j of a reciprocity script, and the joint interplay of all three variables? Neglecting all constant parameters for the moment, we can write:

$$\mathbf{o_i} * \mathbf{a_j} > 1 - S / \mathbf{U}$$

where S is the constant derived from (C/p). Obviously, the threshold depends on all three parameters at the same time, and whether a single parameter change "tips over" the threshold balance crucially depends on the specification of all other parameter values. That is to say, the model predicts two- and three-way interactions between U, o_i and a_j. In a statistical model, we would have to include not only main effects U, o_i, and a_j, but also interaction terms ($U*o_i$), ($U*a_j$), ($a_j*o_{i'}$) and the three-way interaction ($U*a_j*o_i$). But what is the predicted sign of these effects?

Note that each experimental condition and parameter constellation will provide for some range [0,a*]==A_{low} and [a*,1]==A_{high} of a_j in which the activation weight AW($A_{high}|o_i$,U) > RHS, and AW($A_{low}|o_i$,U) < RHS, that is, the threshold defined by the right-hand side (RHS) is reached for A_{high} and it is not reached for A_{low}. The threshold-value a* can (but need not) be different for all four experimental conditions (thus, denote each a* with A1-A4). What is more, an accessibility-value larger than a* may *not* be sufficient to "tip over" the threshold balance because the remaining constant parameters have an unfavorable specification. We need to ask whether a change in o_i or U is *sufficient* to induce a shift from the rational to the automatic mode in either range of a_j, whether both parameters are jointly *necessary* to induce this shift, or whether their joint effect is *not* sufficient. The threshold condition may even remain unfulfilled when both factors support the automatic mode, because the constant parameters (opportunity p, link l_i, cost of reflection C, temporary script accessibility $a_{j|i}$) push the balance into an unfavorable region where the effect of a parameter change disappears. All these possibilities have to be taken care of when thinking about the potential outcomes in each experimental condition (see Kroneberg 2006a, 2011b).

The following table summarizes the hypothesized impact of the experimental treatments on the threshold value along with chronic accessibility ranges A_{low} and A_{high}. It shows all effects of a parameter change on the left-hand-side (displaying the activation weight, AW) and the right-hand-side (RHS) of the mode-selection threshold, along with the resulting outcome, which is either the rational or the automatic mode. Every experimental condition or a shift in accessibility can "tip over" the threshold balance and trigger the rational *or* the automatic mode. For the incentive treatment, assume that U take the values $U_{low}<U_{high}$; for the context treatment, assume that $o_{neutral}<o_{coop}$ (table 1):

Table 1: Experimental treatments and **c**hanges in the mode selection threshold

	RHS= 1- S/U / AW= a_j*o_i	Incentives U_{low} RHS decreases	$AW= a_j*o_i$	Incentives U_{high} RHS increases
Context Neutral $o_{neutral}$	$A1_{low} * o_{neutral}$	1. > → automatic	$A3_{low} * o_{neutral}$	9. > → automatic
		2. < → rational		10. < → rational
	$A1_{high} * o_{neutral}$ LHS increases	3. > → automatic	$A3_{high} * o_{neutral}$ LHS increases	11. > → automatic
		4. < → rational		12. < → rational
Context Cooperative o_{coop}	$A2_{low} * o_{coop}$ LHS increases	5. > → automatic	$A4_{low} * o_{coop}$ LHS increases	13. > → automatic
		6. < → rational		14. < → rational
	$A2_{high} * o_{coop}$ LHS increases	7. > → automatic	$A4_{high} * o_{coop}$ LHS increases	15. > → automatic
		8. < → rational		16. < → rational

Note: Outcomes of mode-selection are presented as a function of experimental conditions U (initial endowments) and o (framing condition) in conjunction with chronic accessibility a*.

With the help of simple logic, we can exclude all combinations from the total of $2^8 = 256$ different outcome patterns which are contradictive and therefore not feasible. For example, it is not possible that (1/4/6/8) is reached simultaneously, because the automatic mode was selected in the most unfavorable condition (1) already, and the activation weight on the left-hand-side can never decrease in conditions (4), (6) or (8); thus the rational mode can never become selected given that (1) is true. In this way, we can logically exclude the pairwise combinations 1/4, 1/6, 1/8, 3/8, 5/8, 9/12, 9/14, 9/16, 11/14, 11/16, 13/16, 2/9, 4/11, 6/13, 8/15, 4/9, 8/13, which restricts the potential interaction patterns to a number of 17 admissible patterns:

1. 1,3,5,7,9,11,13,15 (always automatic)
2. 1,3,5,7,10,11,13,15
3. 1,3,5,7,10,12,13,15
4. 1,3,5,7,10,12,14,15
5. 1,3,5,7,10,12,14,16
6. 2,3,5,7,10,11,13,15
7. 2,3,5,7,10,12,13,15
8. 2,3,5,7,10,12,14,15
9. 2,3,5,7,10,12,14,16
10. 2,3,6,7,10,12,14,15
11. 2,3,6,7,10,12,14,16
12. 2,4,5,7,10,12,13,15
13. 2,4,5,7,10,12,14,15
14. 2,4,5,7,10,12,14,16
15. 2,4,6,7,10,12,14,15
16. 2,4,6,7,10,12,14,16
17. 2,4,6,8,10,12,14,16 (always rational)

In the following, I will present a graphical solution to the problem of predicting interaction patterns. The results will be demonstrated using pattern number 16, which is selected here at random for presentational purposes only. The principal setup and procedures are similar for any other admissible interaction pattern. A full list of all graphical solutions to the derived patterns can be obtained from the author on request. Given that pattern 16 is statistically observed, we can update the table to show all mode-selection contingencies (table 2):

Table 2: Predicted interaction pattern #16 and mode selection contingencies

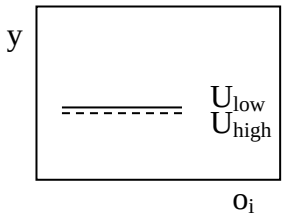
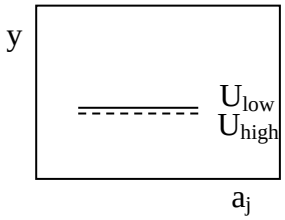
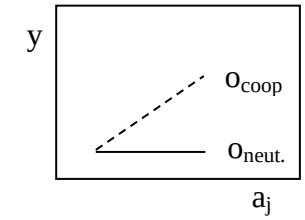
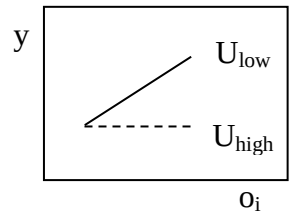
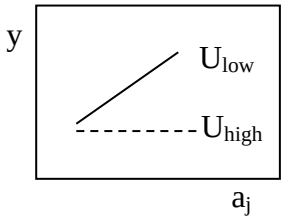
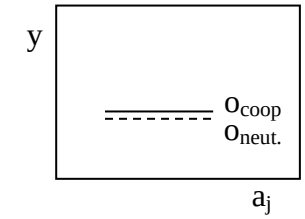
	RHS= 1- S/U / AW= a_j*o_i	Incentives U_{low} RHS decreases	AW= a_j*o_i	Incentives U_{high} RHS increases
Context Neutral $o_{neutral}$	$A1_{low}$ * $o_{neutral}$	1. > → automatic	$A3_{low}$ * $o_{neutral}$	9. > → automatic
		2. < → rational		10. < → rational
	$A1_{high}$ * $o_{neutral}$ LHS increases	3. > → automatic	$A3_{high}$ * $o_{neutral}$ LHS increases	11. > → automatic
		4. < → rational		12. < → rational
Context Cooperative o_{coop}	$A2_{low}$ * o_{coop} LHS increases	5. > → automatic	$A4_{low}$ * o_{coop} LHS increases	13. > → automatic
		6. < → rational		14. < → rational
	$A2_{high}$ * o_{coop} LHS increases	7. > → automatic	$A4_{high}$ * o_{coop} LHS increases	15. > → automatic
		8. < → rational		16. < → rational

How would the conditional effects on the level of trust (*reltrust*) in each experimental condition look like, given that this pattern is observed? From the table we can see that:

(1) The conditional effect (CE) of a_j is zero in the low incentive / neutral context condition
(2) The CE of a_j is positive in the low incentive / coop. context condition (cells 6 to 7)
(3) The CE of a_j is zero whenever incentives are high (neutral and cooperative context)
(4) The CE of the context o_i is positive in the low incentive / A_{high} condition (cells 4 to 7)
(5) The CE of the context o_i is zero in all other conditions (high incentives or A_{low})
(6) The CE of incentives U is negative in the coop. context / A_{high} condition (cells 7 to 16)
(7) The CE of U is zero in all other conditions (neutral context or A_{low})

Using these conditional effects, we can graphically pin down all outcomes and interactions. First, fix one variable at one level. In a graph, let the x-axis display the level of the second variable, using the y-axis to graph *reltrust*, using the bridge hypotheses proposed above as a guide. The CE of the second variable can be graphed for each level of the third variable. In the following table, each graph refers to another way of displaying the information that can be obtained from table 2, holding constant one variable and varying the remaining two each time. In this way, derive the predicted sign of all two-way interactions for each experimental condition and factorial combination. The sign of the three-way interaction can be inferred from observing all two-way interactions and their common direction of change (see table 3):

Table 3: Predicted interaction pattern #16 and *reltrust*

Constant: A_{low}  CE o_i = zero CE U = zero Interaction Effect = zero	Constant: $o_{neutral}$  CE U = zero CE a = zero Interaction Effect = zero	Constant: U_{low}  CE o_i = zero if A_{low} = positive if A_{high} CE $a_{j\|i}$ = zero if $o_{neutral}$ = positive if o_{coop} Interaction Effect = positive
Constant: A_{high}  CE o_i = positive if U_{low} = zero if U_{high} CE U = zero if $o_{neutral}$ = negative if o_{coop} Interaction Effect = negative	Constant: o_{coop}  CE U = zero if A_{low} = negative if A_{high} CE a_j = positive if U_{low} = zero if U_{high} Interaction Effect = negative	Constant: U_{high}  CE o_i = zero CE a_j = zero Interaction Effect = zero
IE Change: zero → negative	IE Change: zero → negative	IE Change: positive → zero

a_j	≥ 0	
U	≤ 0	
o_i	≥ 0	
U x o_i	≤ 0	
a_j x U	≤ 0	
a_j x o_i	≥ 0	
a_j x U x o_i	**< 0**	(inferred from the IE changes presented above)

Note: CE = Conditional Effect; y= predicted level of *reltrust*; A_{low} (A_{high}) = level of chronic script accessibility aj below (above) a*; U= incentive treatment, varying on two levels U_{low} , U_{high}; o_i = context treatment, varying on two levele $o_{neutral}$ and o_{coop}; The table displays the conditional effects of remaining parameters, holding one parameter constant at a time. The constant parameter is indicated in the top of each box. The CE can be inferred from the contingency table, as presented above. The two-way interactions can be inferred from graphing each CE within each condition.

This procedure can be repeated for all remaining interaction patterns. The resulting set of interaction patterns has been presented in section 6.3.3 already, it is repeated here for completeness (see table 4 below):

Table 4: Predicted interaction patterns for *reltrust*

	Predicted Interaction Patterns (Main- and Interaction Effects)																
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a_j	0	≥ 0	≥ 0	≥ 0	$= 0$	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	0
U	0	≤ 0	≤ 0	≤ 0	< 0	$= 0$	$= 0$	≤ 0	≤ 0	≤ 0	< 0	$= 0$	≤ 0	≤ 0	$= 0$	≤ 0	0
o_i	0	≥ 0	≥ 0	≥ 0	$= 0$	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	$= 0$	> 0	≥ 0	≥ 0	≥ 0	≥ 0	0
$U \cdot o_i$	0	≥ 0	> 0	≥ 0	$= 0$	$= 0$	≥ 0	≥ 0	≥ 0	≥ 0	$= 0$	$= 0$	≤ 0	< 0	$= 0$	≤ 0	0
$a_i \cdot U$	0	≥ 0	$= 0$	≥ 0	$= 0$	$= 0$	≤ 0	≤ 0	≤ 0	≤ 0	$= 0$	$= 0$	≥ 0	$= 0$	$= 0$	≤ 0	0
$a_i \cdot o_i$	0	≤ 0	$= 0$	≥ 0	$= 0$	< 0	≤ 0	≤ 0	≤ 0	≥ 0	$= 0$	$= 0$	≥ 0	$= 0$	> 0	≥ 0	0
$a_i \cdot o_i \cdot U$	0	≤ 0	$= 0$	> 0	$= 0$	$= 0$	> 0	> 0	> 0	> 0	$= 0$	$= 0$	> 0	$= 0$	$= 0$	< 0	0

Note: The table presents predicted interaction patterns between chronic script accessibility a_j, situational cues o_i and motivation U to predict transfer decisions (*reltrust*) in the investment game