TRUST FROM A TRAIT PERSPECTIVE: A THEORETICAL FRAMEWORK AND EMPIRICAL TEST

ISABEL THIELMANN

M. Sc.

Inaugural dissertation

submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences

in the Graduate School of Economic and Social Sciences

at the University of Mannheim

Dean of the Faculty of Social Sciences:	Prof. Dr. Michael Diehl
Academic Director of the CDSS:	Prof. Dr. Thomas Bräuninger
Thesis Reviewers:	Prof. Dr. Edgar Erdfelder
	Dr. Jochen E. Gebauer
Thesis Defense:	December 8th, 2015
Defense Committee:	Prof. Dr. Dagmar Stahlberg
	Prof. Dr. Edgar Erdfelder
	Dr. Jochen E. Gebauer

CONTENTS

ABSTRACT

Corresponding to the pivotal role of trust for all kinds of social interactions and interpersonal relationships, trust has been the target of abundant research across scientific disciplines. However, an integration of the huge literature is currently missing – thus hampering a common understanding of trust and a synthesis between the fields. Furthermore, from a psychological perspective, there is an insufficient understanding of the (basic) personality traits accounting for individual variation in trust. Therefore, the overall objective of this thesis is to bridge the gap between different lines of trust research and to uncover the dispositional determinants of trust. To that end, a behavioral view on trust is adopted, defining trust in terms of a risky choice to depend on another. Based on a broad review of the literature, a theoretical framework is distilled, identifying the situational features and personality characteristics underlying trust. Specifically, trust is considered to be a function of (1) attitudes toward risky prospects (risk and loss aversion), (2) trustworthiness expectations, and (3) betrayal sensitivity. These determinants are, in turn, rooted in different traits (i.e., anxiety/fear, trustworthiness, and forgiveness) which can be localized in the space defined by basic personality models. Here, the HEXACO model provides particularly clear-cut hypotheses on the basic traits driving trust, including a unique factor for each of the proposed (specific) trait determinants.

Building on this reasoning, the empirical part of this thesis presents first evidence on the link between the HEXACO dimensions and trust. As a starting point, the focus was on the Honesty-Humility factor, representing the unique feature of the HEXACO model compared to more established models of personality (e.g., the Five-Factor Model). In line with the proposed theoretical framework, two sets of studies provided support for a social projection path from trait trustworthiness to trustworthiness expectations. Specifically, high levels of Honesty-Humility predicted more optimistic trustworthiness expectations and – as necessitated by social projection – were also positively linked to trustworthy behavior. As such, the findings not only identify a trait source of trust, but also clarify the dispositional determinants of trustworthiness. Overall, the theoretical framework and empirical evidence presented in this thesis suggests the fruitfulness to take a closer look at trust – from a trait perspective.

1. ARTICLES

This thesis is based on a set of three articles that have been published in peer-reviewed journals. The articles will be discussed – and are appended to this work – in the order in which they are listed below. This order mirrors the chronological order of development, demonstrating the deductive approach adopted in this thesis (i.e., from theoretical framework to empirical evidence). It does, however, not perfectly correspond to the time of publication (which is why the second article is already cited in the first). In general, note that in summarizing the articles throughout this thesis, I will refrain from reiterating any details of the particular articles which can be found in the articles as appended to this work.

(1) Thielmann, I., & Hilbig, B. E. (2015b). Trust: An integrative review from a person-situation perspective. *Review of General Psychology, 19*(3), 249-277. doi: 10.1037/gpr0000046

(2) Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality*, *50*(3), 61-65. doi: 10.1016/j.jrp.2014.03.006

(3) Thielmann, I., & Hilbig, B. E. (2015a). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin, 41*(11), 1523-1536. doi: 10.1177/0146167215600530

Note that four more articles related to the topic at hand (i.e., pro-social behavior in economic games) have been published during the development of this thesis (see below). However, given that these articles leave trust and trustworthiness aside, they do not constitute an integral part of this thesis. Nonetheless, they will be referred to in the main text whenever appropriate.

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S., & Zettler, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality, 55*, 46-50. doi: 10.1016/j.jrp.2014.12.004

Hilbig, B. E., Thielmann, I., Wührl, J., & Zettler, I. (2015). From Honesty-Humility to fair behavior – Benevolence or a (blind) fairness norm? *Personality and Individual Differences, 80*, 91-95. doi: 10.1016/j.paid.2015.02.017

Thielmann, I., Böhm, R., & Hilbig, B. E. (2015). Different games for different motives: Comment on Haesevoets, Folmer, and Van Hiel (2015). *European Journal of Personality, 29*(4), 506-508. doi: 10.1002/per.2007

Thielmann, I., Hilbig, B. E., & Niedtfeld, I. (2014). Willing to give but not to forgive: Borderline personality features and cooperative behavior. *Journal of Personality Disorders, 28*(6), 778-795. doi: 10.1521/pedi_2014_28_135

"Trust is the glue of life. It's the most essential ingredient
in effective communication. It's the foundational
principle that holds all relationships."

Stephen Covey, Roger Merrill, & Rebecca Merrill

2. INTRODUCTION AND THEORETICAL BACKGROUND

Trust as "the glue of life" (Covey, Merrill, & Merrill, 1994, p. 203) constitutes an essential pillar of all kinds of social interactions and relationships, even among strangers. It is thus unsurprising that trust has been the target of abundant scientific research for several decades, in psychology and beyond. In what follows, I will first introduce the two most prominent conceptualizations of trust as used in the diverse literature and outline how I will refer to the term "trust" throughout this thesis. Building on this, I will then elaborate on the theoretical foundation of this thesis (Article 1), including a detailed definition of trust as well as a summary of its determinants and their trait basis (if applicable). The subsequent empirical part of this thesis will be concerned with empirically testing one of the theoretically implied paths from personality to trust (Article 2) – which also necessitates considering the counterpart of trust, namely trustworthiness (Article 3). Finally, I will outline some empirical and theoretical implications and draw general conclusions.

Before going into detail, however, some clarification seems in order at the outset. Specifically, note that this thesis will exclusively be concerned with *interpersonal* trust (i.e., trust in individuals, not in groups or organizations) and trust in others' *intentions* (i.e., not trust in others' expertise or skill, respectively). For example, a trusting act as discussed herein (i.e., interpersonal trust in intentions) would be to borrow money to a person from whom you expect not to be exploited. A similar instance would be to ask a stranger in the train to have an eye on your luggage while you visit the restrooms (expecting that the stranger will refrain from steeling your luggage). By contrast, interpersonal trust in another's expertise – as when trusting in a medicine that she will be successful with your knee surgery – or trust in an organization – as when trusting the government that they care for peace and order in your country – represent different types of trust that will not be addressed in this thesis.

2.1 CONCEPTUALIZATIONS OF TRUST

As noted above, there is a long tradition of trust research in various scientific disciplines. Corresponding to this diversity of research, the term "trust" has been used and conceptualized in different ways. Basically, two broad conceptualizations of trust can be distinguished: (1) trust as an attitude and (2) trust as a behavior. Note that, historically, both conceptualizations have been developed almost in parallel. That is, none of the conceptualizations

directly builds on the other, and both understandings coexist until today. In what follows, I will briefly summarize the essence of these two conceptualizations and outline how I will conceptualize trust in the remainder of this thesis.

(1) Trust as an attitude

Understanding trust in terms of an *attitude* implies that trust is basically cognitive in nature. Broadly speaking, an attitude denotes "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Chaiken, 1993, p. 1). In a situation of trust, this "entity" – or object of evaluation, respectively – might either refer to the trusted party (i.e., the *trustee*) or to the trusting act itself. Correspondingly, two types of trust attitudes can be distinguished: trust as a mere expectation, on the one hand, and trust as an intention, on the other hand.

In psychology, the view that trust basically represents an *expectation* regarding another's trustworthiness has first been discussed by Julian Rotter (1967, 1971). In his seminal work, Rotter described interpersonal trust in terms of "an expectancy […] that the word, promise, verbal or written statement of another individual or group can be relied upon" (Rotter, 1967, p. 651). As a measurement instrument, Rotter (1967) developed the Interpersonal Trust Scale, thereby setting the cornerstone for the assessment of trust via self-report questionnaires (for similar measures see, e.g., Johnson-George & Swap, 1982; Larzelere & Huston, 1980; Yamagishi, 1986). Until today, Rotter's understanding of trust as an expectation has been very well received and is still shared by several scholars (e.g., Ben-Ner & Halldorsson, 2010; Gambetta, 1988; Robinson, 1996; Sitkin & Roth, 1993).

Besides defining trust in terms of an expectation, another attitudinal view of trust implies that trust basically mirrors an *intention*, thus rendering the trusting act itself the entity of evaluation. According to this view, trust denotes the "willingness to be vulnerable to another party" (Mayer, Davis, & Schoorman, 1995, p. 726) based on some expectation regarding another's trustworthiness (see also, e.g., McKnight, Cummings, & Chervany, 1998; Rousseau, Sitkin, Burt, & Camerer, 1998). The trust attitude as defined along these lines hence involves a vulnerability aspect beyond mere expectations. However, note that trust is still distinguished from consequential (risk-taking) behavior. That is, according to the idea that trust reflects an intention of making oneself vulnerable – and an attitudinal understanding of trust more generally – "one does not need to risk anything in order to trust" (Mayer et al., 1995, p. 724).

(2) Trust as a behavior

Unlike the view that trust is different from a consequential risk-taking action, understanding trust in terms of a *behavior* implies that trust basically mirrors a risky choice. Morton Deutsch (1958, 1962) was the first researcher to adopt this view in psychology. In his pioneering work, Deutsch conceptualized trust as a "choice of an ambiguous path" (Deutsch, 1973, p. 149), based upon some expectation that a desirable event will occur. In essence, Deutsch hence considered those aspects emphasized in attitudinal views of trust (i.e., expectations and vulnerability, as summarized above) as prerequisites of trust – which, in itself, is conceptualized in terms of a decision to depend on another. As such, Deutsch operationalized trust in game-theoretic (social dilemma) paradigms in which an individual's outcome – once she has trusted – is entirely contingent upon another's action (Deutsch, 1960, 1973; see also Swinth, 1967; Wrightsman, 1966, for similar approaches).

Until today, several scholars have adopted an equivalent behavioral understanding of trust, describing trust as a choice to rely or depend on another – and thus as a special case of risk-taking[1] (e.g., Boon & Holmes, 1991; Coleman, 1990; Fehr, 2009; Kee & Knox, 1970; Riker, 1974; Yamagishi, Kanazawa, Mashima, & Terai, 2005). The development of the Trust Game in economics (Berg, Dickhaut, & McCabe, 1995) has finally heavily stimulated corresponding research on trust behavior in the past 20 years (for a meta-analytic review see Johnson & Mislin, 2011). In particular, the Trust Game seeks to provide a unique measure of trust given that trust is considered the primary motive driving behavior (cf. Berg et al., 1995).[2] Taken together, conceptualizing trust in terms of a behavior hence implies that trust is consequential in that it corresponds to a risky decision to entrust a personal resource to another.

[1] Although Deutsch explicitly separated trust behavior from risk-taking or gambling, respectively, he implicitly incorporated risk-taking as an aspect of trust. Specifically, Deutsch considered that the trusting choice is inevitably linked to the possibility of a loss – which essentially implies a risk as conceptualized in this thesis and elsewhere (e.g., Das & Teng, 2004; Orbell, 1993; see also Footnote 3).

[2] In the Trust Game, the so-called *trustor* is endowed with a certain amount of money from which she can entrust any proportion to her interaction partner, the *trustee*. The entrusted amount is typically tripled by the experimenter before passed on to the trustee who can then decide how much to return to the trustor. Commonly, the amount entrusted by the trustor is considered a measure of trust behavior whereas the amount returned by the trustee is considered a measure of trustworthy behavior. The Trust Game is hence sought to provide a purer measure of trust compared to the social dilemma games originally used by Deutsch (which heavily confound trust with cooperation; e.g., Yamagishi et al., 2005). However, note that other motives have also been discussed as drivers underlying trust decisions in the Trust Game (e.g., social welfare maximization, fairness; cf. Thielmann, Böhm, & Hilbig, 2015). Thus, it is questionable whether the Trust Game indeed overcomes the limitations associated with previous game-theoretic measures of trust.

To sum up, two broad conceptualizations of trust can be distinguished: trust as an attitude versus trust as a behavior. Whereas the former clearly separates trust from risk-taking, the latter implies that trust inherently involves a risk and corresponding action. In fact, it has been noted elsewhere (and will be detailed in the next section) that risk constitutes a key characteristic of trust given that "the twin of trust is betrayal" (Dunn, 1988, p. 81) – thus rendering a loss resulting from trust inevitably possible. Conceptualizing trust in terms of an attitude that separates trust from risk-taking misses out on this vital aspect. Furthermore, understanding trust in terms of an attitude leaves trust without any consequences – a view that is completely at odds with corresponding economic approaches emphasizing the importance of incentives and outcomes (cf. Lopes, 1994). Hence, a purely attitudinal understanding of trust might endanger the fruitful collaboration between psychologists and economists which is still in its infancy (Handgraaf & Fred van Raaij, 2005) after years of "suspicion and distaste" (Lopes, 1994, p. 198). In this regard, it seems also worth mentioning that, originally, "psychology is the science of behavior" (Baumeister, Vohs, & Funder, 2007, p. 396) – a fact that has increasingly fallen into oblivion since the cognitive revolution in the late 20th century and will likewise be undermined by a purely attitudinal conceptualization of trust.

Based on this reasoning, I will adopt a behavioral understanding of trust in this thesis. However, as sketched above, this does certainly not imply that attitudes are generally excluded from the concept of trust. Rather, a behavioral view allows an integration of attitudes as prerequisites of trust behavior. Against this background, in the next section I will provide a detailed definition of trust – which I have, admittedly, left aside so far – and clarify the underlying determinants of trust as distilled based on a broad review of the trust literature.

2.2 DEFINING FEATURES AND DETERMINANTS OF TRUST

Thielmann, I., & Hilbig, B. E. (2015b). Trust: An integrative review from a person-situation perspective. *Review of General Psychology, 19*(3), 249-277. doi: 10.1037/gpr0000046

As mentioned previously, trust is at the heart of various social interactions, even among strangers. Hence, trust has been the target of abundant research across different fields. The main goal in Thielmann and Hilbig (2015b; Article 1) was to review and integrate this diverse literature into a coherent structure to arrive at a common understanding of trust and its determinants and to consequently bridge the gap between different scientific disciplines. For this

purpose, we first considered it vital to incorporate findings from both personality and social psychology and to profit from their fruitful synthesis (e.g., Fleeson & Noftle, 2008; Funder, 2008). In this regard, we specifically intended to overcome the current neglect of (basic) personality traits as underlying determinants of trust behavior. So far, approaches to trust mainly considered *trust propensity*, that is, a person's "general willingness to trust others" (Mayer et al., 1995, p. 715), as a specific trait to account for the large individual variation in trust (Johnson & Mislin, 2011). In light of the multifaceted nature of trust behavior (as will be detailed below), this approach is apparently highly oversimplified. Moreover, we particularly aimed at integrating the vast literature from economics and related fields – corresponding to the behavioral view on trust adopted herein and the "collaborative potential of psychology and economics" (Handgraaf & Fred van Raaij, 2005, p. 388). So, on the whole, we provided an integrative person-situation framework on trust behavior based on a broad review of the literature from personality psychology, social psychology, and economics, particularly focusing on trust among unknown agents.

In what follows, I will summarize the essence of our review and framework. Although I will refrain from reiterating details and references that can be found in the article as appended to this thesis, the summary of this particular article will be quite detailed given that it provides the core theoretical background of this thesis.

Defining trust

As a first step, we considered it important to provide an *integrative* definition of trust behavior that incorporates different perspectives on trust and, more importantly, captures the key features of trust as proposed in previous work (e.g., Rousseau et al., 1998) and as sketched above: (a) uncertainty and risk[3], (b) expectations, and (c) vulnerability. At first, there is high agreement that trust behavior is necessarily associated with uncertainty and risk. *Uncertainty* implies that, in a situation of trust, the trustor can never have conclusive knowledge about the trustee's trustworthiness and, in turn, the likelihood of trust appreciation versus trust betrayal (e.g., Yamagishi & Yamagishi, 1994). In consequence, trusting inevitably involves

[3] Following the decision-making literature, uncertainty implies that the decision maker is unfamiliar with the odds of gain versus loss. Risk, by contrast, suggest that the decision maker can experience both a gain or a loss, but that the corresponding odds are known (e.g., Ellsberg, 1961; Orbell, 1993). Typically, risk and uncertainty covary given that low uncertainty is associated with low risk and high uncertainty is associated with high risk.

risk (e.g., Coleman, 1990; Das & Teng, 2004) due to the possibility of a loss resulting from potential betrayal. As follows from this inherent uncertainty, on the one hand, and risk, on the other hand, the trustor has to base her decision to trust on *expectations* regarding the other's trustworthiness (e.g., Boon & Holmes, 1991; Deutsch, 1958) and to accept personal *vulnerability* (e.g., Das & Teng, 2004; Mayer et al., 1995), that is, a loss due to misplaced trust.

In sum, we hence define trust behavior as "a risky choice of making oneself dependent on the actions of another in a situation of uncertainty, based upon some expectation of whether the other will act in a benevolent fashion despite an opportunity to betray" (p. 251). As such, our definition offers a straightforward operationalization of trust in terms of a risky dependence choice. However, although we clearly conceptualize trust in terms of a behavior, we nonetheless incorporate those aspects that have been emphasized in attitudinal conceptualizations of trust as summarized above (i.e., expectations and vulnerability).

Determinants of trust

As follows from considering the decision to trust structurally similar to a decision under risk, trust behavior should involve four basic components (according to the decision-making literature on risky choice): the probability of a gain (i.e., trust appreciation), the utility of that gain, the probability of a loss (i.e., trust betrayal), and the utility of that loss (Payne, 1973; Slovic & Lichtenstein, 1968). Based on this notion – and the key characteristics of trust as outlined above (i.e., risk and uncertainty, expectations, and vulnerability) – we distilled three core determinants underlying the decision to trust from the literature: (I) attitudes toward risky prospects (i.e., risk aversion and loss aversion), (II) trustworthiness expectations, and (III) betrayal sensitivity (see Figure 1, for a graphical illustration). Notably, each determinant either refers to the probabilities or to the utilities of the potential outcomes of the trusting choice. In what follows, I will briefly elaborate on these three determinants of trust and further delineate their trait bases (in terms of specific as well as basic personality traits). For the sake of brevity, I will refrain from reiterating evidence supporting the corresponding conjectures which can be found in the article as appended to this thesis.

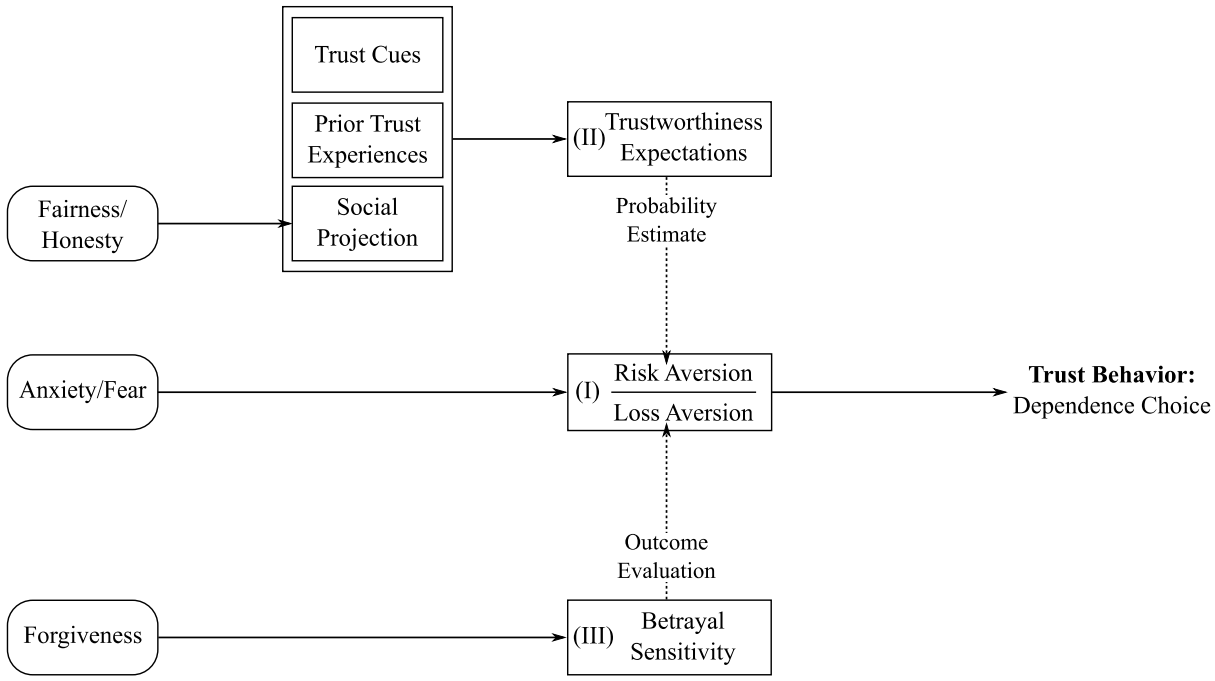

Figure 1. Graphical illustration of the determinants of trust behavior as summarized in Thielmann and Hilbig (2015b). Solid arrows denote causal relationships; dashed arrows denote information input. Personality traits are framed with round boxes. Adapted from "Trust: An integrative review from a person-situation perspective" by Thielmann and Hilbig (2015b). Copyright 2015 by the American Psychological Association.

According to prospect theory (Kahneman & Tversky, 1979, 1984) – and corresponding to the risk and uncertainty aspects of trust – two *attitudes toward risky prospects* (I) should arguably drive the decision to trust: risk aversion and loss aversion. More detailed, *risk aversion* denotes an individual's tendency to prefer a sure (positive) outcome over a potentially higher, yet risky outcome (cf. Glöckner & Hilbig, 2012). Regarding the decision to trust, several scholars hence agree that individuals high in risk aversion should require a high subjective probability of encountering a trustworthy interaction partner (i.e., high probability of a gain) in order to prefer trust (the risky option) over distrust (the safe option). By contrast, for individuals low in risk aversion, a comparably small probability of encountering a trustworthy trustee might suffice to opt for the (risky) trusting choice. *Loss aversion*, in turn, denotes an individual's tendency to evaluate a loss as being more aversive than a gain of comparable magnitude attractive (cf. Bibby & Ferguson, 2011). Concerning the decision to trust, loss aversion might hence influence the trustor's evaluation of the positive utility (i.e., the gain) resulting from trust appreciation in relation to the negative utility (i.e., the loss) resulting from trust betrayal. That is, the higher the loss aversion, the more weight (or severity, respectively)

should be assigned to the potential loss resulting from betrayal, thus rendering trust less likely. Altogether, whereas risk aversion should influence the processing of the (subjective) probabilities of the potential gain versus loss associated with trust, loss aversion should influence the processing of the corresponding utilities. In terms of personality traits, evidence implies that both attitudes toward risky prospects are linked to trait anxiety and fear, respectively.

Besides these two attitudes toward risky prospects, we summarize *trustworthiness expectations* (II) as a second determinant of trust decisions. As sketched above, the inherent uncertainty in a trust situation forces the trustor to infer the trustee's trustworthiness in the absence of conclusive knowledge. Ample evidence suggests that trustors make use of different sources of information to form these trustworthiness expectations, namely trust cues, prior trust experiences, and social projection. *Trust cues* are somewhat observable pieces of evidence available in the environment (cf. lens model; Brunswik, 1952) that can either refer to characteristics of the trustee (i.e., outward appearance, reputation, social category) or to features of the trust situation (i.e., temptation to betray, presence of potential sanctions for untrustworthy behavior). *Prior trust experiences*, in turn, emphasize the assumed learning aspect underlying trust behavior. Specifically, trustors seem to consult their past experiences in similar situations to estimate the probability of trust appreciation versus trust betrayal in a new situation (e.g., Blair & Stout, 2001; Rotter, 1967). Finally, it has been noted that individuals might use their own trustworthiness to infer others' trustworthiness. In particular, such a *social projection* account implies that individuals project their own tendencies onto others, thus expecting others to be as trustworthy as they themselves are (e.g., Krueger & Acevedo, 2005; Krueger, DiDonato, & Freestone, 2012). As such, an individual's trait trustworthiness – that is, her trait fairness or honesty, respectively, depending on the specific trust situation and trust object (e.g., a material good vs. a secret) – should account for inter-individual variation in trust behavior. On the whole, irrespective of which pieces of information a trustor actually uses, the resulting trustworthiness expectations should arguably provide the (subjective) probability input on which risk aversion operates (cf. Figure 1).

Third and finally, we proposed *betrayal sensitivity* (III) as a determinant underlying trust, based on the notion that individuals differentiate between losses resulting from another's betrayal and losses resulting from nature (e.g., bad luck; cf. Rabin, 1993). In particular, various evidence suggests that people have a general tendency to perceive losses resulting from betrayal to be more severe than losses resulting from nature, a phenomenon called *betrayal*

aversion (e.g., Bohnet & Zeckhauser, 2004). Correspondingly, we define betrayal sensitivity in terms of individual differences in betrayal aversion as "an individual's tendency to attribute a greater severity to a loss resulting from human selfishness than to a formally equivalent loss resulting from nature" (p. 261). In other words, the higher the betrayal sensitivity, the more severely an individual should evaluate the negative utility resulting from potential betrayal. Betrayal sensitivity will hence provide the outcome input on which loss aversion operates (cf. Figure 1). In terms of a more basic trait, betrayal sensitivity should arguably be rooted in trait forgiveness.

Trust in models of basic personality structure

As mentioned above, a central goal of our review and framework was to bridge the gap between different lines of research, with a particular focus on integrating approaches and evidence from personality psychology into the trust literature. Therefore, as a next step, we linked the distilled determinants of trust (i.e., attitudes toward risky prospects, trustworthiness expectations, and betrayal sensitivity) and their proposed trait bases (i.e. anxiety and fear, trustworthiness, and forgiveness) to models of basic personality structure – which provide a broad and parsimonious account to the study of individual differences (e.g., Funder, 2001; Ozer & Reise, 1994). In particular, we identified the basic trait dimensions that should, according to the proposed framework, account for individual variation in trust behavior. To this end, we concentrated on two personality models that have mainly been considered in recent research on social behavior: the Five-Factor Model (FFM; Costa & McCrae, 1992; McCrae & Costa, 1985) and the HEXACO model of personality structure (Ashton & Lee, 2007; Lee & Ashton, 2004).

Corresponding to its name, the FFM includes five factors, namely Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. As follows from their theoretical conceptualizations and corresponding evidence, two of these, Neuroticism and Agreeableness, should be relevant for trust to occur. Specifically, it has repeatedly been shown that individual differences in anxiety and fear are accounted for by Neuroticism. Besides, trait trustworthiness – with its components fairness and honesty – seems to be best reflected in the Agreeableness factor of the FFM, which basically involves differences in the motivation to cooperate versus defect (Denissen & Penke, 2008). Finally, forgiveness has consistently been linked to a mixture of said dimensions, namely low levels of Neuroticism and high levels of FFM-

Agreeableness. In total, the FFM would thus account for individual variation in trust behavior through Neuroticism and Agreeableness – a conjecture that has already obtained some (albeit weak) empirical support.

As an extension and slight variation of the FFM, the HEXACO model – with its name derived as an acronym of the factors it constitutes, namely Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to Experience – has recently been established based on lexical studies across diverse languages. In particular, these studies suggest that personality is best described in terms of six rather than five broad trait dimensions (Ashton et al., 2004; Lee & Ashton, 2008).[4] As such, the HEXACO model proposes three factors accounting for the quality of social interactions and pro-social behavior: Honesty-Humility, Emotionality (i.e., the counterpart of FFM-Neuroticism), and Agreeableness (Ashton, Lee, & de Vries, 2014). More detailed, Honesty-Humility and Agreeableness capture complementary aspects of reciprocal altruism (Trivers, 1971), that is, fairness versus forgiveness. Emotionality, in turn, represents tendencies relevant for kin altruism (e.g., empathetic concern, emotional attachment, and harm-avoidance). Notably, each of these three pro-social traits can, at least in theory, be linked to one of the trait aspects that are assumed to underlie trust (according to our framework). First, anxiety and fear are both captured by the Emotionality factor and its corresponding facets. Second, trustworthiness is arguably covered by Honesty-Humility incorporating *sincerity* as well as *fairness* at the facet-level (see Section 3.1 for details and supporting evidence). And third, forgiveness is mirrored in the Agreeableness factor of the HEXACO model (see Section 3.3 for further information). Overall, compared to the FFM, the HEXACO model hence provides more clear-cut hypotheses on the trait determinants of trust behavior, thus allowing for particularly strict empirical testing. However, as evidenced by a recent meta-analysis (Zhao & Smillie, 2015), actual evidence on the link between trust behavior (in the Trust Game) and the HEXACO dimensions is fully missing.

[4] Specifically, three HEXACO dimensions, namely Extraversion, Conscientiousness, and Openness to Experience, are nearly identical in content to their FFM-counterparts. However, for the remaining dimensions, the HEXACO model incorporates considerable changes. Most strikingly, Honesty-Humility is proposed as a new, sixth trait factor, capturing content beyond the FFM. Furthermore, the HEXACO model comprises alternative rotations of Neuroticism (termed Emotionality) and Agreeableness. In consequence, the model offers particularly straightforward theoretical interpretations of the different trait factors (for further details see Ashton et al., 2014).

To sum up, our review and framework provides an integrative summary and coherent structure of the extant trust literature from different fields. In essence, it implies that (a) trust can be regarded as a risky dependence choice, (b) multiple person and situation characteristics may influence the decision to trust, and (c) individual differences in trust are likely rooted in a combination of different personality traits. Hence, it seems heavily oversimplified to assume that one specific trait dimension (such as trust propensity) might sufficiently account for individual variation in trust behavior. That said, a critical reflection is nonetheless in order: First and foremost, our conclusions were almost exclusively based on trust among strangers. It is thus an important quest for future research to clarify whether and how the proposed framework applies or can be extended to other types of trust. Moreover, for several aspects suggested in the framework – especially those related to the proposed (basic) traits underlying trust – evidence is remarkably weak, thus emphasizing the necessity for critical empirical tests in future studies. In this regard, it seems also worthwhile to take specific person-situation-interactions into account which we have, admittedly, left aside from our discussion due to the deficiency of corresponding evidence.

3. BEYOND THEORY

As summarized above, several gaps in the trust literature still exist, thus inevitably rendering our theoretical framework partly speculative in nature. First and foremost, there is only rudimentary knowledge of the trait basis underlying trust behavior. That is, whereas the situational aspects driving trust are well-investigated, this is, in fact, not the case for the potential personality aspects underlying trust. In particular, evidence on the link between trust behavior and basic personality traits is remarkably weak – with a complete lack of evidence on the dimensions of the more recently proposed HEXACO model (cf. Zhao & Smillie, 2015).

Notably, however, the HEXACO model has already proven to allow for an exceptionally fine-grained analysis of pro-social tendencies, especially when evidence on the somewhat broader FFM is inconclusive (cf. Hilbig, Glöckner, & Zettler, 2014). Correspondingly, the HEXACO model also provides particularly clear-cut hypotheses on the trait basis of trust. That is, as detailed above, each theoretically-implied trait aspect of trust behavior (i.e., anxiety/fear, trustworthiness, and forgiveness) is arguably reflected in one specific HEXACO dimension (i.e., Emotionality, Honesty-Humility, and Agreeableness). Hence, it seems particularly promising to investigate individual differences in trust behavior based on the HEXACO traits.

According to this reasoning, the following empirical part of this thesis is concerned with individual differences in trust (and trustworthy) behavior from the perspective of the HEXACO model. Specifically, given that the predictive advantage of the HEXACO model is primarily reducible to the inclusion of Honesty-Humility as a sixth trait dimension – capturing several aspects that are not well accommodated by the FFM (e.g., Ashton & Lee, 2008; Lee, Ogunfowora, & Ashton, 2005) – I considered Honesty-Humility as a promising starting point for empirical testing. Hence, I will particularly focus on Honesty-Humility in what follows and, corresponding to our person-situation framework, on the proposed link between trait trustworthiness (arguably covered by Honesty-Humility) and trustworthiness expectations. As such, the overarching objective of the empirical part of this thesis was to break the first ground on bridging the gap between trust and personality research and to enhance the understanding of the basic traits underlying trust.

3.1 HONESTY-HUMILITY AND TRUSTWORTHINESS EXPECTATIONS: A PATH OF SOCIAL PROJECTION?

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality, 50*(3), 61–65. doi: 10.1016/j.jrp.2014.03.006

In Thielmann and Hilbig (2014; Article 2), our goal was to provide first evidence on the notion that trait trustworthiness (in terms of individual levels in Honesty-Humility) might account for individual variation in trustworthiness expectations through social projection. The theoretical starting point for this conjecture was two-fold: On the one hand, it has been noted on numerous occasions that social projection constitutes a reasonable antecedent of trust behavior (e.g., Krueger et al., 2012; Krueger, Massey, & DiDonato, 2008). That is, individuals might expect others to be as trustworthy as they themselves are – an assumption nicely corresponding to the typically observed positive relation between individuals' willingness to trust and to behave trustworthily (e.g., Colquitt, Scott, & LePine, 2007; Wrightsman, 1966). On the other hand, trait trustworthiness is arguably captured by the HEXACO Honesty-Humility factor. Supporting this notion – and corresponding to the theoretical conceptualization of Honesty-Humility subsuming sincerity, fairness, greed-avoidance, and modesty – a substantial body of evidence has proven Honesty-Humility to be a valid predictor of fairness-related behaviors (e.g., Baumert, Schlösser, & Schmitt, 2013; Hilbig, Thielmann, Hepp, Klein, & Zettler, 2015; Hilbig, Zettler, & Heydasch, 2012) as well as of honesty (Hilbig, Moshagen, & Zettler, 2015; Hilbig & Zettler, 2015) – the two pillars of trustworthiness, as outlined above.

Integrating these two pieces of evidence, our aim was to test the hypothesis that individuals infer others' trustworthiness from their own trait trustworthiness (i.e., their Honesty-Humility levels) by means of social projection. More precisely, we assumed that individuals high in Honesty-Humility will have more optimistic expectations regarding others' trustworthiness than their counterparts low in Honesty-Humility because they themselves are more trustworthy or cooperative, respectively. In a study (N = 244), we assessed individuals' Honesty-Humility levels (Ashton & Lee, 2009) and their expectations regarding another's trustworthiness in the Distrust Game (McEvily, Radzevick, & Weber, 2012) – an alternative to the Trust Game that offers a more straightforward operationalization of trustworthiness expectations (because in the Distrust Game the trustee's behavior is not contingent on the trustor's

behavior). To test the social projection account more conclusively, we additionally included measures of cooperativeness (i.e., a hypothetical Dictator Game; Forsythe, Horowitz, Savin, & Sefton, 1994) and entitlement (i.e., a hypothetical Ultimatum Game; Güth, Schmittberger, & Schwarze, 1982). By this means, we intended to explicitly verify that individuals high in Honesty-Humility expect others to be more trustworthy because they themselves are, rather than because they feel entitled to a good treatment by others.

In line with this hypothesis, we found a positive relation between Honesty-Humility and trustworthiness expectations which was mediated through cooperativeness in the Dictator Game, but not through entitlement in the Ultimatum Game. Hence, the data supported the idea that social projection of one's own trait trustworthiness matters for the formation of trustworthiness expectations. However, it should be acknowledged that the effect of Honesty-Humility on trustworthiness expectations was only small to medium-sized. On the one hand, this might be due to the exclusive reliance on hypothetical games without real interaction partners or incentives involved. On the other hand, individual differences in the tendency to rely on social projection (Krueger & Acevedo, 2007) might have acted as a moderator of said relationship, thus reducing the overall effect size. Furthermore, another shortcoming of our study is that we entirely focused on individuals' trustworthiness expectations, without considering a measure of actual trust behavior. Although (a priori) expectations regarding another's trustworthiness have been shown to reliably drive trust behavior (e.g., Coricelli, Gonzalez Morales, & Mahlstedt, 2006; Kugler, Connolly, & Kausel, 2009; Yamagishi et al., 2013), combining the presented design with a measure of trust behavior might be a worthwhile extension for future research – alongside the use of real interactions involving actual incentives.

3.2 FROM TRUST TO TRUSTWORTHINESS

As another limitation of the above reasoning, the postulated social projection account basically rests on the notion that Honesty-Humility actually explains individual variation in trustworthiness. Although this assumption seems well-justified in light of the empirically demonstrated links between Honesty-Humility and pro-social/honest behavior, direct evidence on said relation is still missing. Thus, for the social projection account to hold, it seems necessary to critically test whether Honesty-Humility indeed explains individual variation in trustworthy behavior, that is, variation in the appreciation of another's trust.

By and large, evidence on the (basic) traits underlying trustworthiness is scarce anyway. First, extant findings on the FFM are inconclusive as, for example, indicated by a weak meta-analytic correlation between FFM-Agreeableness and trustworthiness in the Trust Game (Zhao & Smillie, 2015). Thus, investigating trustworthy behavior from the perspective of the HEXACO model, specifically the Honesty-Humility factor, might be a promising approach to clarify the inconsistent empirical picture (cf. Hilbig et al., 2014). Moreover, prior research is still vague on the specific mechanism underlying trustworthiness that might, in turn, drive the relation between basic personality traits and trustworthy behavior. In particular, it is unclear whether trustworthiness is an expression of positive reciprocity, negative reciprocity, or unconditional kindness (e.g., Ashraf, Bohnet, & Piankov, 2006; Ben-Ner & Halldorsson, 2010; Cox, 2004). Stated differently, evidence is inconclusive on whether trustworthiness mirrors a conditional behavior that is contingent on the level of the trustor's prior trust (thus reflecting reciprocity) or whether it mirrors an unconditional behavior in that it involves a relatively stable reaction that does not perfectly correspond to the trustor's previous behavior (thus reflecting kindness).

Notably, illuminating the nature and dispositional determinants of trustworthiness might not only enhance the understanding of trustworthy behavior as such, but also of the rationale underlying trust. That is, given that trust is only profitable and reasonable if the corresponding reaction is trustworthy to some extent (cf. Evans & Krueger, 2011; Hardin, 1996), there should be no trust at all in the absence of trustworthiness. In other words, trustworthiness *begets* trust (e.g., Hardin, 2002; Tullberg, 2008). Thus, knowledge about the (trait) determinants driving trustworthiness will provide basic insights into the factors that foster trust in the long run. Based on this reasoning, in Article 3 we (Thielmann and Hilbig, 2015a) dissected the trait determinants underlying trustworthy behavior by means of the HEXACO model of personality. Importantly, this approach did not only allow us to illuminate the specific mechanism driving trustworthiness, but also to critically test whether Honesty-Humility indeed explains individual differences in trustworthy behavior – the basic conjecture underlying the proposed social projection path to trustworthiness expectations.

3.3 DISSECTING THE TRAIT DETERMINANTS OF TRUSTWORTHINESS

Thielmann, I., & Hilbig, B. E. (2015a). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin, 41*(11), 1523-1536. doi: 10.1177/0146167215600530

As mentioned previously, there is currently only a rudimentary understanding of the nature and trait determinants of trustworthy behavior. On the one hand, evidence on the basic traits underlying trustworthiness is restricted to the FFM, with inconclusive findings overall. On the other hand, it is still unclear whether trustworthy behavior is an expression of positive reciprocity, negative reciprocity, or unconditional kindness. In Thielmann and Hilbig (2015a; Article 3), the overall objective was hence to dissect the potential dispositional factors underlying trustworthiness based on the HEXACO model of personality.

As an advantage, the HEXACO model distinguishes between different cooperative tendencies given that it does, most prominently, specifically capture *active cooperativeness* (i.e., fairness or non-exploitation) in the factor of Honesty-Humility and *reactive cooperativeness* (i.e., forgiveness or non-retaliation) in the factor of Agreeableness – as empirically supported by a pattern of double-dissociation (Hilbig, Zettler, Leist, & Heydasch, 2013; Thielmann, Hilbig, & Niedtfeld, 2014). Corresponding to this distinction – and of particular interest for the issue at hand – Honesty-Humility has been explicitly linked to unconditional kindness (e.g., Hilbig, Thielmann, Wührl, & Zettler, 2015) and positive reciprocity (e.g., Ackermann, Fleiß, & Murphy, in press) whereas HEXACO-Agreeableness has been specifically associated with negative reciprocity (e.g., Ackermann et al., in press). Consequently, for each of the proposed mechanisms to trustworthiness, the HEXACO model provides detailed and specific predictions on the to-be-expected link between trustworthiness and Honesty-Humility or HEXACO-Agreeableness, respectively, and the potential moderating role of the level of prior trust shown by the trustor. This allows for strict empirical tests of the competing mechanisms. In detail, if trustworthiness is an expression of unconditional kindness, there should be a positive link between trustworthiness and Honesty-Humility, irrespective of the level of prior trust. By contrast, the relation between trustworthiness and Honesty-Humility should be moderated by the level of trust if trustworthiness mirrors positive reciprocity. Finally, if negative reciprocity is the driving mechanism, trustworthiness should be negatively linked to HEXACO-Agreeableness, again as a function of the level of prior trust.

In a set of three web-based studies, participants acted as the trustees in the Trust Game (Berg et al., 1995) who either specified their trustworthiness in reaction to all potential trust levels a trustor might show (Studies 1 and 2; N = 108 and N = 118, respectively) or to one specific trust level only (Study 3; N = 177). Besides, we assessed the HEXACO dimensions (Ashton & Lee, 2009) and, in two studies, additionally the FFM dimensions (Borkenau & Ostendorf, 2008). Overall, all three studies revealed a positive relation between Honesty-Humility and trustworthy behavior. Importantly, this relationship was unaffected by the trustor's level of prior trust. HEXACO-Agreeableness, by contrast, indicated no association with trustworthiness whatsoever. Altogether, the pattern of results was hence compatible with the predictions as derived from the unconditional kindness mechanism, but contradicted the positive and negative reciprocity mechanisms. This conclusion was further underpinned by the results of Study 3, showing that, for example, the positive relation between Honesty-Humility and trustworthy behavior was mediated by an unconditional kindness justification, but not by a positive or negative reciprocity justification of trustees' return decisions. As such, the results also reconciled the inconsistent evidence on the link between FFM-Agreeableness and trustworthy behavior. In particular, those aspects predicting unconditional kindness are only marginally represented in FFM-Agreeableness, thus producing a small overall effect.

In summary, the reported studies enhance the understanding of trustworthiness in terms of its underlying personality traits. Importantly, they support the assumption that Honesty-Humility accounts for individual variation in trustworthy behavior, thereby validating and complementing the above reasoning underlying the social projection path from trait trustworthiness (in terms of Honesty-Humility) to trustworthiness expectations. Nonetheless, some critical remarks are worth mentioning. First, the web-based procedure might have insufficiently triggered reciprocity due to the minimal interpersonal contact between interaction partners (i.e., trustors and trustees). A replication of the above results in a lab-based setting with a higher degree of interpersonal contact is thus desirable. Also, the findings are entirely based on a specific trust situation (i.e., the Trust Game) and a specific type of trust relationship (i.e., trust among strangers), thus questioning the generalizability of conclusions. For example, it might be the case that in close relationships, trustworthiness is more a matter of (positive/negative) reciprocity than of unconditional kindness because individuals might expect close others to be trustful toward them. Future studies might critically test this conjecture and look for potential moderators of the relation between personality traits and trustworthiness.

4. CONCLUSIONS AND OUTLOOK

The theoretical framework and empirical evidence reported on in this thesis revealed insights into individual differences in interpersonal trust and trustworthiness from a (basic) trait perspective. In Thielmann and Hilbig (2015b; Article 1), we provided a broad review and integrative summary of the diverse trust literature, based on which we distilled a person-situation framework on trust behavior. As such, we introduced a theoretical foundation for the study of individual differences in trust. Building on this, in Thielmann and Hilbig (2014; Article 2) we presented corresponding empirical support for one of the proposed sources of individual variation in trust: an individual's own trustworthiness. Specifically, we showed that individuals apparently use their trait trustworthiness (operationalized via individuals' Honesty-Humility levels) to infer others' trustworthiness – reflecting a path of social projection underlying trustworthiness expectations. However, given that this reasoning rested on the (untested) notion that Honesty-Humility accounts for trustworthy behavior, in Thielmann and Hilbig (2015a; Article 3) we finally shifted our focus from trust to the personality aspects predicting trustworthiness. In line with the postulated social projection account, we provided first empirical evidence on the predictive power of Honesty-Humility for trustworthy behavior. At the same time, we clarified the dispositional determinants underlying trustworthiness and demonstrated that trustworthy behavior is basically a question of unconditional kindness.

Overall, the work presented in this thesis sets a starting point for an enhanced understanding of individual differences in trust (and trustworthiness) – an issue that has gained insufficient attention in prior research. Nevertheless, there are still several gaps and open quests that should be addressed in future work. Most strikingly, evidence on the two additional paths from personality to trust behavior as proposed in our person-situation framework (i.e., from trait anxiety/fear via attitudes toward risky prospects to trust behavior and from trait forgiveness via betrayal sensitivity to trust behavior; cf. Figure 1) is inconclusive or even missing so far. Moreover, transferring the theoretical assumptions and empirical evidence presented in this thesis to trust and trustworthiness in other types of relationships (e.g., friends or romantic partners) seems a desirable extension to be targeted in future research. In general, it is my hope that the work presented herein, especially the provided theoretical framework on trust, will stimulate systematic research on the (basic) trait sources underlying individual differences in trust (and trustworthy) behavior among several types of trust relationships.

Beyond revealing future research avenues, the work presented in this thesis also has important theoretical implications for the study of trust. In particular, it first seems advisable to conceptualize trust in terms of a behavior – corresponding to the "original goal [of psychology] of being the science of behavior" (Baumeister et al., 2007, p. 400). Notably, such a behavioral view on trust allowed us to take a particularly fine-grained look at the different determinants of trust. At the same time, a behavioral view might also foster the fruitful symbiosis between different areas of trust research, most notably psychology and economics (Handgraaf & Fred van Raaij, 2005). Besides, within psychology, research on trust might profit from a synthesis between social and personality psychology (e.g., Fleeson & Noftle, 2008; Funder, 2008). In this regard, I hope to have elucidated the specific usefulness to take into account the actors' personality characteristics beyond situational features (as primarily focused on in prior research) to get a complete understanding of the occurrence of trust among strangers. Recent research on cooperation has already demonstrated the fruitfulness of considering (basic) personality traits to account for individual differences in cooperative behavior (e.g., Zhao & Smillie, 2015) – an approach that has recently even found its way into economics (Handgraaf & Fred van Raaij, 2005). In the area of trust research, it may likewise be the time to increasingly integrate a trait perspective to illuminate the emergence of trust behavior. By this means, we might achieve a more holistic understanding of the puzzle of trust as the "foundational principle that holds all relationships" (Covey et al., 1994, p. 203).

5. REFERENCES

Ackermann, K. A., Fleiß, J., & Murphy, R. O. (in press). Reciprocity as an individual difference. *Journal of Conflict Resolution*. doi: 10.1177/0022002714541854

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193–208. doi: 10.1007/s10683-006-9122-4

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. doi: 10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008). The prediction of Honesty-Humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality*, *42*(5), 1216–1228. doi: 10.1016/j.jrp.2008.03.006

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. doi: 10.1080/00223890902935878

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139–152. doi: 10.1177/1088868314523838

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., … De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. doi: 10.1037/0022-3514.86.2.356

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. doi: 10.1111/j.1745-6916.2007.00051.x

Baumert, A., Schlösser, T. M., & Schmitt, M. (2013). Economic games: A performance-based assessment of fairness and altruism. *European Journal of Psychological Assessment*. doi: 10.1027/1015-5759/a000183

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79. doi: 10.1016/j.joep.2009.10.001

Berg, J., Dickhaut, J., & McCabe, K. A. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. doi: 10.1006/game.1995.1027

Bibby, P. A., & Ferguson, E. (2011). The ability to process emotional information predicts loss aversion. *Personality and Individual Differences*, *51*(3), 263–266. doi: 10.1016/j.paid.2010.05.001

Blair, M. M., & Stout, L. A. (2001). Trust, trustworthiness, and the behavioral foundations of corporate law. *University of Pennsylvania Law Review*, *149*(6), 1735–1810. doi: 10.2307/3312898

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, *55*(4), 467–484. doi: 10.1016/j.jebo.2003.11.004

Boon, S. D., & Holmes, J. G. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In R. A. Hinde & J. Groebel (Eds.), *Cooperation and prosocial behavior* (pp. 190–211). Cambridge, England: Cambridge University Press.

Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI). Manual (2., neu normierte und vollständig überarbeitete Auflage).* Göttingen: Hogrefe.

Brunswik, E. (1952). *The conceptual framework of psychology.* Oxford: University Chicago Press.

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA US: Belknap Press.

Colquitt, J. A., Scott, B. A., & LePine, J. A. (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. doi: 10.1037/0021-9010.92.4.909

Coricelli, G., Gonzalez Morales, L., & Mahlstedt, A. (2006). The investment game with asymmetric information. *Metroeconomica*, *57*(1), 13–30.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and NEO Five–Factor Inventory (NEO–FFI) professional manual.* Odessa, FL US: Psychological Assessment Resources.

Covey, S., Merrill, R. A., & Merrill, R. R. (1994). *First things first: To live, to love, to learn, to leave a legacy*. New York, US: Simon and Schuster.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281. doi: 10.1016/S0899-8256(03)00119-2

Das, T. K., & Teng, B.-S. (2004). The risk-based view of trust: A conceptual framework. *Journal of Business and Psychology*, *19*(1), 85–116. doi: 10.1023/B:JOBU.0000040274.23551.1b

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor Model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*(5), 1285–1302. doi: 10.1016/j.jrp.2008.04.002

Deutsch, M. (1958). Trust and suspicion. *The Journal of Conflict Resolution*, *2*(4), 265–279. doi: 10.2307/172886

Deutsch, M. (1960). Trust, trustworthiness, and the F scale. *The Journal of Abnormal and Social Psychology*, *61*(1), 138–140. doi: 10.1037/h0046501

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1962.* (pp. 275–320). Oxford: University Nebraska Press.

Deutsch, M. (1973). *The resolution of conflict: constructive and destructive processes*. *The Carl Hovland memorial lectures*. New Haven: Yale University Press.

Dunn, J. (1988). Trust and political agency. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 73–93). Oxford: Basil Blackwell.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Orlando, FL, US: Harcourt Brace Jovanovich College Publishers.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, *75*, 643–669.

Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk and perspective-taking. *Journal of Experimental Social Psychology*, *47*(1), 171–177. doi: 10.1016/j.jesp.2010.08.007

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, *7*(2-3), 235–266.

Fleeson, W., & Noftle, E. (2008). The end of the person–situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, *2*(4), 1667–1684. doi: 10.1111/j.1751-9004.2008.00122.x

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369.

Funder, D. C. (2001). Personality. *Annual Review of Psychology*, *52*, 197–221. doi: 10.1146/annurev.psych.52.1.197

Funder, D. C. (2008). Persons, situations, and person-situation interactions. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research (3rd ed.).* (pp. 568–580). New York, NY US: Guilford Press.

Gambetta, D. (1988). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–237). Oxford: Basil Blackwell.

Glöckner, A., & Hilbig, B. E. (2012). Risk is relative: Risk aversion yields cooperation rather than defection in cooperation-friendly environments. *Psychonomic Bulletin & Review*, *19*(3), 546–553. doi: 10.3758/s13423-012-0224-z

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388. doi: 10.1016/0167-2681(82)90011-7

Handgraaf, M. J. J., & Fred van Raaij, W. (2005). Fear and loathing no more: The emergence of collaboration between economists and psychologists. *Journal of Economic Psychology*, *26*(3), 387–391. doi: 10.1016/j.joep.2004.12.003

Hardin, R. (1996). Trustworthiness. *Ethics*, *107*(1), 26–42.

Hardin, R. (2002). *Trust and trustworthiness.* New York, NY: Russell Sage Foundation.

Hilbig, B. E., Glöckner, A., & Zettler, I. (2014). Personality and pro-social behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology*, *107*(3), 529–539. doi: 10.1037/a0036074

Hilbig, B. E., Moshagen, M., & Zettler, I. (2015). Truth will out: Linking personality, morality, and honesty through indirect questioning. *Social Psychological and Personality Science*, *6*(2), 140–147. doi: 10.1177/1948550614553640

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S., & Zettler, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality*. doi: 10.1016/j.jrp.2014.12.004

Hilbig, B. E., Thielmann, I., Wührl, J., & Zettler, I. (2015). From Honesty–Humility to fair behavior – Benevolence or a (blind) fairness norm? *Personality and Individual Differences*, *80*, 91–95. doi: 10.1016/j.paid.2015.02.017

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*, 72–88. doi: 10.1016/j.jrp.2015.04.003

Hilbig, B. E., Zettler, I., & Heydasch, T. (2012). Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional Honesty-Humility. *European Journal of Personality*, *26*(3), 245–254. doi: 10.1002/per.830

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598–603. doi: 10.1016/j.paid.2012.11.008

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. doi: 10.1016/j.joep.2011.05.007

Johnson-George, C., & Swap, W. C. (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. doi: 10.1037/0022-3514.43.6.1306

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341.

Kee, H. W., & Knox, R. E. (1970). Conceptual and methodological considerations in the study of trust and suspicion. *Journal of Conflict Resolution*, *14*(3), 357–366. doi: 10.1177/002200277001400307

Krueger, J. I., & Acevedo, M. (2005). Social projection and the psychology of choice. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment.* (pp. 17–41). New York, NY: Psychology Press.

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning. *The American Journal of Psychology*, *120*(4), 593–618.

Krueger, J. I., DiDonato, T. E., & Freestone, D. (2012). Social projection can solve social dilemmas. *Psychological Inquiry*, *23*(1), 1–27. doi: 10.1080/1047840X.2012.641167

Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence to social norms. *Negotiation and Conflict Management Research*, *1*(1), 31–52. doi: 10.1111/j.1750-4716.2007.00003.x

Kugler, T., Connolly, T., & Kausel, E. E. (2009). The effect of consequential thinking on trust game behavior. *Journal of Behavioral Decision Making*, *22*(2), 101–119. doi: 10.1002/bdm.614

Larzelere, R. E., & Huston, T. L. (1980). The Dyadic Trust Scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family*, *42*(3), 595–604. doi: 10.2307/351903

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. doi: 10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2008). The HEXACO personality factors in the indigenous personality lexicons of English and 11 other languages. *Journal of Personality*, *76*(5), 1001–1053. doi: 10.1111/j.1467-6494.2008.00512.x

Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality traits beyond the Big Five: Are they within the HEXACO space? *Journal of Personality*, *73*(5), 1437–1463. doi: 10.1111/j.1467-6494.2005.00354.x

Lopes, L. L. (1994). Psychology and economics: Perspectives on risk, cooperation, and the marketplace. *Annual Review of Psychology*, *45*, 197–227. doi: 10.1146/annurev.ps.45.020194.001213

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. doi: 10.2307/258792

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, *49*(3), 710–721. doi: 10.1037/0022-3514.49.3.710

McEvily, B., Radzevick, J. R., & Weber, R. A. (2012). Whom do you distrust and how much does it cost? An experiment on the measurement of trust. *Games and Economic Behavior*, *74*(1), 285–298. doi: 10.1016/j.geb.2011.06.011

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *The Academy of Management Review*, *23*(3), 473–490. doi: 10.2307/259290

Orbell, J. M. (1993). Hamlet and the psychology of rational choice under uncertainty. *Rationality and Society*, *5*(1), 127–140. doi: 10.1177/1043463193005001010

Ozer, D. J., & Reise, S. P. (1994). Personality assessment. *Annual Review of Psychology*, *45*, 357–388. doi: 10.1146/annurev.ps.45.020194.002041

Payne, J. W. (1973). Alternative approaches to decision making under risk: Moments versus risk dimensions. *Psychological Bulletin*, *80*(6), 439–453. doi: 10.1037/h0035260

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*(5), 1281–1302.

Riker, W. H. (1974). The nature of trust. In J. T. Tedeschi (Ed.), *Perspectives on Social Power* (pp. 63–81). Chicago: Aldine.

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, *41*(4), 574–599. doi: 10.2307/2393868

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*(5), 443–452. doi: 10.1037/h0031464

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. F. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. doi: 10.5465/AMR.1998.926617

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science*, *4*(3), 367–392. doi: 10.1287/orsc.4.3.367

Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology*, *78*(3, Pt.2), 1–18. doi: 10.1037/h0026468

Swinth, R. L. (1967). The establishment of the trust relationship. *The Journal of Conflict Resolution*, *11*(3), 335–344. doi: 10.2307/172600

Thielmann, I., Böhm, R., & Hilbig, B. E. (2015). Different games for different motives: Comment on Haesevoets, Folmer, and Van Hiel (2015). *European Journal of Personality*, *29*(4), 506–508. doi: 10.1002/per.2007

Thielmann, I., Hilbig, B. E., & Niedtfeld, I. (2014). Willing to give but not to forgive: Borderline personality features and cooperative behavior. *Journal of Personality Disorders*, *28*(6), 778–795. doi: 10.1521/pedi_2014_28_135

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Tullberg, J. (2008). Trust—The importance of trustfulness versus trustworthiness. *The Journal of Socio-Economics*, *37*(5), 2059–2071. doi: 10.1016/j.socec.2007.10.004

Wrightsman, L. S. (1966). Personality and attitudinal correlates of trusting and trustworthy behaviors in a two-person game. *Journal of Personality and Social Psychology*, *4*(3), 328–332. doi: 10.1037/h0023655

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, *51*(1), 110–116. doi: 10.1037/0022-3514.51.1.110

Yamagishi, T., Kanazawa, S., Mashima, R., & Terai, S. (2005). Separating trust from cooperation in a dynamic relationship: Prisoner's dilemma with variable dependence. *Rationality and Society*, *17*(3), 275–308. doi: 10.1177/1043463105055463

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., … Simunovic, D. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations

of pro-sociality. *Organizational Behavior and Human Decision Processes*, *120*(2), 260–271. doi: 10.1016/j.obhdp.2012.06.002

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 129–166. doi: 10.1007/BF02249397

Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review*, *19*(3), 277–302. doi: 10.1177/1088868314553709

ACKNOWLEDGMENTS

It would have been impossible to accomplish this thesis without the endless support from several people whom I wish to thank at this point.

First, I would like to express my heartfelt gratitude to Benjamin Hilbig, my mentor and advisor. His support during the last years was exceptional, both on a scientific and a personal level. Benni always had a watchful eye over what I was doing and provided guidance whenever needed. I owe him that I feel well-prepared for the diverse challenges that come along with working in academia. I am happy that I have placed my trust in him, and do very much look forward to our future collaboration.

Also, particular thanks go to my highly esteemed colleagues and friends Ann-Katrin Wesslein, Robert Böhm, Dorothee Mischkowski, and Friederike Theobald for their continuous emotional support as well as their helpful comments on previous versions of this thesis. Besides, I would like to warmly thank my colleagues from the Cognitive Psychology Lab at the University of Koblenz-Landau as well as my fellow grad students at the CDSS. They always had a sympathetic ear and were on hand with help and advice.

Finally, I wish to express my heartfelt thanks to my beloved family. My parents brought me up to an independent and self-assured individual and always supported me in what I was doing. I am really thankful for their trust in me and that they always were so proud. My deepest thanks lastly go to Patrick Beschorner for his endless patience, sympathy, and love. He backed me up whenever I was assailed by doubts, and placed all his faith and trust in me. It is a true gift having found one's best friend for life.

STATEMENT OF ORIGINALITY

I hereby declare that I am the sole author of this thesis and have made use of no other sources than those cited in this work.

Mannheim, October 2015 Isabel Thielmann

CO-AUTHOR'S STATEMENT

It is hereby confirmed that the following articles included in this thesis were primarily conceived and written by Isabel Thielmann, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim.

Thielmann, I., & Hilbig, B. E. (2015b). Trust: An integrative review from a person-situation perspective. *Review of General Psychology, 19*(3), 249-277. doi: 10.1037/gpr0000046

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality*, *50*(3), 61–65. doi: 10.1016/j.jrp.2014.03.006

Thielmann, I., & Hilbig, B. E. (2015a). The traits one can trust: Dissecting reciprocity and kindness as determinants of trustworthy behavior. *Personality and Social Psychology Bulletin*, *41*(11), 1523-1536. doi: 10.1177/0146167215600530

I sign this statement to the effect that Isabel Thielmann is credited as the primary source of ideas and the main author of all three articles.

Mannheim, October 2015 Prof. Benjamin E. Hilbig, Ph. D.

APPENDIX: COPIES OF ARTICLES

Review of General Psychology
2015, Vol. 19, No. 3, 249–277

1089-2680/15/$12.00 http://dx.doi.org/10.1037/gpr0000046

Trust: An Integrative Review From a Person–Situation Perspective

Isabel Thielmann
University of Koblenz-Landau and University of Mannheim

Benjamin E. Hilbig
University of Koblenz-Landau

Trust is a key aspect of various social interactions. Correspondingly, trust has been heavily studied across different scientific disciplines. However, an integration of the diverse research and literature is still missing. Addressing this issue, we review several hundred articles on interpersonal trust among strangers and integrate them into a coherent framework, explaining trust behavior among unfamiliar agents based on an interaction between situational features and distinct personality characteristics. Understanding trust as a decision under risk, we distill 3 core components of trust behavior from the extant literature: attitudes toward risky prospects (i.e., risk aversion and loss aversion), trustworthiness expectations, and betrayal sensitivity. Each of these refers to a distinct set of causal determinants, including personality characteristics (anxiety/fear, trustworthiness, and forgiveness) which can be localized in the space defined by models of basic personality structure (e.g., the Five-Factor Model and the HEXACO model of personality). In sum, the review contributes to the understanding of trust behavior by linking and integrating the findings from various fields of trust research. Additionally, it provides fruitful directions and implications for future research.

Keywords: betrayal sensitivity, interpersonal trust behavior among strangers, personality traits, risk/loss aversion, trustworthiness expectations

We're never so vulnerable than when we trust someone—but paradoxically, if we cannot trust, neither can we find love or joy.
—Walter Anderson

The importance of trust for all areas of human interaction can probably not be overemphasized (cf. Yamagishi, 2011). As Rotter (1971) summarized, "the entire fabric of our day-to-day living, of our social order, rests on trust" (p. 443). In other words, "perhaps there is no single variable which so thoroughly influences interpersonal and group behaviour as does trust" (Golembiewski & McConkie, 1975, p. 131). Whether in close relationships or in interactions with strangers, we trust on numerous occasions throughout our lives. For example, we trust our partner not to cheat on us, we confide in friends with personal or intimate problems, we employ others to take care of our children, we ask strangers on the train to watch our luggage while we visit the restrooms, we buy second-hand cars hoping to not end up with a lemon, and we purchase goods online without actually knowing the seller or testing the product beforehand. Based on this omnipresence of trust, it is unsurprising that trust is arguably one of the most heavily studied constructs across all social and economic sciences (e.g., psychology, economics, political science, sociology, law) and beyond.

This article was published Online First July 20, 2015.

Isabel Thielmann, Department of Psychology, University of Koblenz-Landau and Center for Doctoral Studies in Social and Behavioral Sciences, University of Mannheim; Benjamin E. Hilbig, Department of Psychology, University of Koblenz-Landau.

Correspondence concerning this article should be addressed to Isabel Thielmann, Department of Psychology, University of Koblenz-Landau, Fortstraβe 7, 76829 Landau, Germany. E-mail: thielmann@uni-landau.de

Traditionally, one type of trust seems particularly relevant, namely trust among strangers—also referred to as *initial trust* (McKnight, Cummings, & Chervany, 1998). Today, the growing e-commerce sector increasingly forces people to trust strangers when purchasing products online. Besides, a variety of social interactions in every-day life require trust in unknown others, thus precluding the long-term establishment of trusting relationships (cf. Bohnet & Zeckhauser, 2004; Dunning, Fetchenhauer, & Schlösser, 2012; Dunning & Fetchenhauer, 2010; Gill, Boies, Finegan, & McNally, 2005). From an economic (game-theoretic) perspective—assuming that humans are rational utility maximizers primarily motivated by self-interest and, by implication, mainly untrustworthy—it seems mostly irrational to trust strangers (cf. Evans & Krueger, 2009; J. M. Weber, Malhotra, & Murnighan, 2005). However, as the examples above demonstrate, people do trust unknown others. Indeed, trust among strangers represents a behavioral tendency that, once developed, remains stable across the life span (Sutter & Kocher, 2007). Based on this apparent discrepancy between rationality and observable behavior, the current work aims at uncovering the determinants underlying this seemingly irrational, yet common and indeed often necessary type of social behavior. In particular, we review and, more importantly, aim to integrate the interdisciplinary literature on trust among strangers, thus providing a coherent summary of this phenomenon.

Within the varied research, different approaches have been put forward to explain trust behavior in general, and trust among strangers in particular. To name a few prominent examples, the literature comprises game-theoretic approaches to trust (e.g., Cabon-Dhersin & Ramani, 2007; Macy & Skvoretz, 1998), approaches focusing on specific aspects of trust (e.g., risk; Das & Teng, 2004) or specific occurrences of trust (e.g., trust in groups; Meyerson, Weick, & Kramer, 1996), and approaches focusing on

particular areas of trust (e.g., organizational trust; Mayer, Davis, & Schoorman, 1995; McKnight et al., 1998). Though all of these approaches come with different foci, most of them share one core feature: Trust is considered to be the result of an interaction between features of the specific trust situation and individual characteristics of the trusting person (i.e., the *trustor*). That is, on the one hand, individuals are commonly assumed to condition their trust on situation-specific variables (e.g., another's trustworthiness), suggesting an intraindividual variability of trust behavior across situations. On the other hand, people are assumed to have a stable inclination toward (dis)trusting, suggesting a person-specific aspect underlying trust and an intraindividual consistency of trust behavior across situations (Fleeson & Leicht, 2006).

Especially regarding this latter point—the individual differences in trust and the personality characteristics responsible for this variation—existing approaches are notably underspecified. That is, existing trait-based accounts mainly refer to *trust propensity* as the underlying personality dimension of trust, defined as an individual's "general willingness to trust others" (Mayer et al., 1995, p. 715) or "one's personal tendency to believe in others' trustworthiness" (Das & Teng, 2004, p. 109). However, as these definitions clearly demonstrate, assigning trust propensity scores to individuals who differ in trust behavior is a mere redescription of observed behavior, rather than an explanation in terms of underlying personality traits (cf. Hilbig, Zettler, & Heydasch, 2012). In terms of theoretical progress, such redescriptions are not particularly useful (Gigerenzer, 1998), thus signaling the need for a more basic trait-based understanding of individual differences in trust. In addition, the implication that trust behavior can be reduced to a single specific personality trait may well be an oversimplification given the apparent complexity of trust behavior (as will be detailed below).

To overcome such limitations, we present an integrative review of the literature on trust among strangers and incorporate the research from different disciplines into a coherent person-situation framework in the sense of a descriptive taxonomy. In this framework, we specify both the situational and the personality aspects underlying trust behavior. Specifically, we summarize theoretical and empirical contributions from behavioral economics, social psychology, and personality psychology—thus bridging the gap between different lines of research. As such, we conclude that (a) trust behavior is a function of multiple situation and person characteristics and (b) individual differences in trust behavior result from a combination of relatively distinct personality traits. At the same time, we present an up-to-date review of the trust literature, particularly considering the various findings put forward in the past decade of trust research. In this regard, we also aim at pointing to the gaps in the trust literature that we hope can be bridged by future research. In general, we focus on interpersonal trust (rather than trust in and among groups or organizations) and trust in intentions (rather than capabilities; cf. Snijders & Keren, 2001). Also, we approach trust from a behavioral perspective, following the growing body of research operationalizing trust through behavior in economic games.[1] Note that this behavioral perspective also corresponds to recent research on conformity (as first promoted by Asch, 1956), which proposes that disagreeing with better informed others—and thus disclosing one's ignorance—is another expression of trust (particularly in others' good-will; Hodges, Meagher, Norton, McBain, & Sroubek, 2014; Hodges, 2014).

The article will be structured as follows: First, we review existing conceptualizations of trust so as to derive an integrative definition of trust behavior—in close adherence to the proposed standards for generating construct definitions (Gilliam & Voss, 2013). This definition is designed to overcome potential limitations of prior definitions and, more importantly, mirrors the essence of the framework we put forward to integrate the extant literature. Based on the definition, we then examine the situational and individual determinants of trust behavior, including their potential interactions, and combine them into a coherent person-situation framework. At this point, we will start off by summarizing the gist of the proposed framework before elaborating on each single component of the framework in detail. As part of this detailed view, we also link the specified personality determinants underlying trust behavior to well-established models of basic personality structure. Finally, we discuss the implications of the presented framework for theory and research and elaborate how it can easily be extended to apply to other types of trust (beyond trust among strangers).

Defining Trust

As sketched above, the concept of trust has been approached from various perspectives and different scientific disciplines. Correspondingly, a substantial number of definitions and conceptualizations of trust have been proposed (for overviews see, e.g., Blomqvist, 1997; Das & Teng, 2004; Hosmer, 1995; McKnight & Chervany, 2001). Despite this diversity, however, scholars from different fields do agree on the basic components of trust (Rousseau, Sitkin, Burt, & Camerer, 1998) which are essentially captured in one of the most heavily cited definitions, describing 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" (Mayer et al., 1995, p. 712). According to this definition—and in line with the common understanding of trust—trusting (a) implies *uncertainty* and *risk*,[2]

[1] Specifically, research on trust behavior has mainly considered the *trust game* (Camerer & Weigelt, 1988; Dasgupta, 1988; Kreps, 1990), also known as the *investment game* (Berg, Dickhaut, & McCabe, 1995; for a meta-analytic review see Johnson & Mislin, 2011). In this game, a trustor decides how to divide an endowment between herself and the so-called *trustee*. The amount of monetary units the trustor is willing to transfer is usually tripled and added to the trustee's earnings. As a response, the trustee can then transfer any amount back to the trustor. In general, the amount sent by the trustor is considered an indicator of trust whereas the amount returned by the trustee is considered an indicator of trustworthiness. Corresponding to this interpretation, investments in the trust game have been shown to be mostly interpreted in terms of risk, faith in others (i.e., beliefs in others' trustworthiness), trust, and gambling (Dunning et al., 2012; see also Dunning et al., 2014).

[2] Note that, in decision making research, *uncertainty* refers to the fact that a decision maker is unfamiliar with the odds of gain versus loss, whereas *risk* refers to the fact that both a gain or a loss can occur (neither is certain) and the decision maker is familiar with the corresponding odds (e.g., Ellsberg, 1961; Orbell, 1993). Thus, in a trust situation, uncertainty exists because the trustor is unfamiliar with the probabilities of gain (i.e., trust appreciation) versus loss (i.e., trust betrayal). Concurrently, risk exists because of the trustee's opportunity to betray and the corresponding possibility for the trustor to experience a loss. In general, uncertainty and risk typically covary given that high certainty will be associated with low risk and low certainty will be associated with high risk, respectively (cf. Yates & Stone, 1992).

given the absence of control on the part of the trustor, (b) is based on an *expectation* that the interaction partner (the trustee) will act in the trustor's interest (i.e., in a benevolent fashion), and (c) requires accepting personal *vulnerability* in terms of potential betrayal. In the following, we will elaborate on these three basic components of trust (see Table 1 for primary references).

To begin with, various approaches imply that trust is inevitably accompanied by uncertainty and risk (e.g., Boon & Holmes, 1991; Das & Teng, 2004; see also Table 1). Given the absence of information about the trustee's intention, a trustor will never be able to conclusively predict the trustee's behavior (Corcos, Pannequin, & Bourgeois-Gironde, 2012; Kramer, 2010; Yamagishi & Yamagishi, 1994). Hence, "it is in situations in which social uncertainty is large that trust is needed" (Yamagishi, 2011, p. 11). As a consequence of this uncertainty, trusting is necessarily accompanied by the possibility of a loss, representing the source of risk associated with trusting (cf. Rousseau et al., 1998).

Similarly following from this uncertainty, a trustor has to rely on her expectation about the trustee's trustworthiness in her decision to trust (e.g., Deutsch, 1958; Yamagishi & Yamagishi, 1994; see also Table 1). That is, before deciding whether it is reasonable to trust or not, a trustor has to estimate the probabilities with which a trustee might honor versus betray her trust. From a rational perspective, a trustor should only trust if it is sufficiently likely that the trustee reacts in a benevolent and favorable fashion; otherwise, trusting would be self-destructive (Evans & Krueger, 2009). Supporting this reasoning, findings show that (a priori) expectations about the trustworthiness of potential interaction partners determine individuals' actual decision to trust in trust games (Bigoni, Bortolotti, Casari, & Gambetta, 2013; Coricelli, Gonzalez Morales, & Mahlstedt, 2006; Deutsch, 1960; Fetchenhauer & Dunning, 2009; Holm & Danielson, 2005; Kugler, Connolly, & Kausel, 2009; Vollan, 2011; Vyrastekova & Garikipati, 2005; Yamagishi et al., 2013), thus pointing to the importance expectations might play for trust to occur.

However, even if a trustor has a somewhat optimistic expectation about the trustee's trustworthiness, she can never be certain that the trustee will actually honor her trust. Always and by definition, betrayal remains "the twin of trust" (J. Dunn, 1988, p. 81)—and thus possible. In essence, betrayal implies a loss for the trustor such that she is worse off after having trusted. Therefore, trusting is necessarily associated with vulnerability toward the trustee's actions (e.g., Mayer et al., 1995; Rousseau et al., 1998; see also Table 1). Vulnerability refers to the severity of adverse outcomes or potential losses (cf. Das & Teng, 2004). It exists whenever the trustee has an incentive to betray the trustor for personal gain (cf. Malhotra, 2004). As this is typically the case in the trust situation, trusting requires taking the risk of a potential loss or, stated differently, accepting vulnerability toward the (potentially betraying) actions of another (e.g., Evans & Krueger, 2011; Ross & LaCroix, 1996; see also Table 1).

Reflecting this vulnerability aspect of trust behavior, trusting has frequently been referred to as a choice of relying or depending on another (e.g., Giffin, 1967; Johnson-George & Swap, 1982; see also Table 1). In a situation of uncertainty, choosing to depend on another necessarily requires accepting personal vulnerability (cf. Boon & Holmes, 1991). For example, entrusting one's child to a babysitter (e.g., to gain an evening out with friends) means to depend on the babysitter's trustworthiness and to accept the possibility that something bad could happen to the child if the babysitter does not take her job seriously. Similarly, asking a stranger on the train to keep an eye on one's luggage means to depend on the stranger's trustworthiness and to accept the possibility that the stranger might steal the luggage. Finally, openly disagreeing with a better informed other and exposing one's ignorance means to depend on another's willingness to appreciate the awkwardness of one's position and to accept the possibility of becoming ostracized (cf. Hodges, 2014). In line with these considerations, a recent meta-analysis identified social dependence as one of the most agreed upon conditions of trust (Balliet & Van Lange, 2012).

In summary, we define interpersonal trust as

> a risky choice of making oneself dependent on the actions of another in a situation of uncertainty, based upon some expectation of whether the other will act in a benevolent fashion despite an opportunity to betray.

In line with previous conceptualizations of trust, this definition includes all core components identified in earlier research (i.e., uncertainty, risk, expectation, and vulnerability toward betrayal), but extends these definitions by incorporating the diverse perspectives found in different scientific disciplines: First, we explicitly distinguish between uncertainty as a feature of the trust situation (i.e., the absence of conclusive information about the trustee's trustworthiness) and risk as a feature of the trusting action (i.e., accepting the possibility of a loss arising from the other's opportunity to betray). Second, we incorporate the idea of separating trust cognitions from trust behavior (e.g., Baron, 1998; Das & Teng, 2004; Fehr, 2009; Hardin, 2002; Kee & Knox, 1970; McKnight et al., 1998; Rotenberg, 2010; Sapienza, Toldra Simats, & Zingales, 2013) by clearly differentiating between expectation (i.e., cognition) and risky dependence choice (i.e., behavior)—which corresponds to recent developments in trust research to the effect that trust behavior and expectations are separately assessed via economic games (e.g., Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014; Yamagishi et al., 2013). Finally, this definition entails a straightforward operationalization of trust behavior for future research, namely that the willingness to depend on another is a direct indicator of the willingness to trust (cf. Table 1). Hence, unlike prior definitions based on which trust was difficult to operationalize (see, e.g., the conceptualizations by Mayer et al., 1995, defining trust as "the willingness [. . .] to be vulnerable", p. 712, or by Rousseau et al., 1998, defining trust as "a psychological state", p. 395), the present definition renders trust directly observable and thus easily quantifiable.[3]

Integrating the Determinants of Trust Behavior

Based on the above definition of trust behavior, we distill three central determinants from the extant literature that should drive the

[3] In general, this conceptualization of trust behavior is in line with a unidimensional view of trust, defining trust based on a continuum ranging from distrust (low values) to trust (high values). In contrast, according to a two-dimensional view, trust and distrust are distinct (albeit related) constructs which can coexist as a result of separate underlying continua ranging from low to high trust and low to high distrust, respectively (e.g., Hill & O'Hara, 2006; Lewicki et al., 1998; McKnight & Chervany, 2001; Sitkin & Roth, 1993).

Table 1
Defining Attributes of Trust with Primary References, Corresponding Determinants (as Included in the Person–Situation Framework), and Implications for Theory and Research

Defining attributes of trust	Primary references	Related determinant(s)	Implications
Trust is inherently accompanied by uncertainty and risk.	(Boon & Holmes, 1991; Camerer, 2003; Coleman, 1990; Das & Teng, 2004; Evans & Krueger, 2011; Gambetta, 1988; Johnson-George & Swap, 1982; Kramer, 2010; Luhmann, 1988; McKnight & Chervany, 2001; Rempel, Holmes, & Zanna, 1985; Rousseau et al., 1998; Schlenker, Helm, & Tedeschi, 1973; B. H. Sheppard & Sherman, 1998; Snijders & Keren, 1999; J. M. Weber et al., 2005; Yamagishi & Yamagishi, 1994)	Risk aversion Loss aversion	Trust behavior can be conceptualized as a decision under risk, involving integration of probabilities and outcomes. Thus, the decision to trust incorporates *two* attitudes toward risky prospects—risk aversion and loss aversion—which should be treated as distinct (albeit complementary) aspects. Inconsistent evidence on the link between trust and risk aversion does not necessarily imply a minor role of risk aversion for trust because it might be attributable to challenges associated with the assessment of risk aversion. It is important to consider the heterogeneity and domain-specificity of risk aversion and thus to carefully select appropriate measures in future studies. Inconsistent evidence on the link between trust and risk aversion might also be attributable to specific interactions between the different trust determinants. These should be addressed in future research. Risk-seeking and low loss aversion might explain individuals' willingness to trust despite expecting a relatively low probability of trust appreciation. The lack of any evidence on the influence of loss aversion on trust behavior should be addressed empirically. Theoretical integration with the vast literature on (cognitive processes of) risky choice is necessary.
Trust is based on expectations about another's trustworthiness.	(Bacharach & Gambetta, 2001; Boon & Holmes, 1991; Deutsch, 1958; Gambetta, 1988; Hardin, 2002; Lewicki, McAllister, & Bies, 1998; McKnight et al., 1998; Pruitt & Kimmel, 1977; Robinson, 1996; Sapienza et al., 2013; Six et al., 2010; Yamagishi & Yamagishi, 1994)	Trustworthiness expectations, based on three sources of information: (a) trust cues (b) prior trust experiences (c) social projection	Trust is not an expectation per se, but expectations are a determinant of trust behavior. Depending on the amount and validity of available information and how it is integrated to form a judgment, trustworthiness expectations are more or less accurate. It should be addressed how individuals search for and integrate information about another's trustworthiness to come up with a judgment. Theoretical integration with the vast literature on judgment processes is needed. The potential influence of prior trust experiences on trust behavior implies that it might be prudent to take into account whether individuals have participated in previous studies using similar paradigms.
Trust requires accepting vulnerability due to the potential betrayal by another.	(Baier, 1986; Bohnet & Zeckhauser, 2004; J. Dunn, 1988; Dunning & Fetchenhauer, 2010, 2011; Evans & Krueger, 2011; Fehr, 2009; Hong & Bohnet, 2007; Malhotra, 2004; Mayer et al., 1995; Ross & LaCroix, 1996; Rousseau et al., 1998; Zand, 1972)	Betrayal sensitivity	Individuals might differ in their weighting of (objectively comparable) losses resulting from another's betrayal versus bad luck. This tendency might in turn affect one's general willingness to trust. Low betrayal sensitivity might explain individuals' willingness to trust despite expecting a relatively low probability of trust appreciation. Research on betrayal sensitivity as a potential source of interindividual variation in trust behavior is needed.

Table 1 (*continued*)

Defining attributes of trust	Primary references	Related determinant(s)	Implications
Trust behavior refers to a choice of depending (or relying) on another.	(Baier, 1986; Boon & Holmes, 1991; Colquitt et al., 2007; Currall & Judge, 1995; Giffin, 1967; Hosmer, 1995; James, 2002; Johnson-George & Swap, 1982; McKnight & Chervany, 2001; McKnight et al., 1998; Moorman, Zaltman, & Deshpande, 1992; Rotenberg, 2010; Schlenker et al., 1973; B. H. Sheppard & Sherman, 1998; Yamagishi et al., 2005)		Trust should be conceptualized as a choice or decision, respectively, rather than as a state or trait (or the like). Similarly, trust behavior has to be distinguished from trust cognitions (i.e., expectations). The choice to depend on another represents a straightforward operationalization of trust behavior in future research. Integration with the decision-making literature is needed.

decision of an individual *X* to (dis)trust an individual *Y* in a situation *Z*: (I) attitudes toward risky prospects (i.e., risk aversion and loss aversion), (II) trustworthiness expectations, and (III) betrayal sensitivity (for a graphical illustration see Figure 1). Before we will discuss these three broad determinants in more detail (and establish links to basic personality traits where appropriate), we provide a brief summary of the gist of the framework in the following. Note that the initial focus of this summary (as well as the more elaborated review thereafter) will be on the component displayed in the center of Figure 1 (i.e., risk aversion/loss aversion) because all other determinants are integrated into a behavioral response at this point. However, although risk aversion and loss aversion hence form the "centerpiece" of this overarching framework, we do not mean to imply that they play a superior role compared to the other determinants. As will be outlined below, it might even be the case under certain circumstances that one particular determinant outweighs the others, thus primarily driving the decision to trust.

Attitudes toward risky prospects (I) capture the risk and uncertainty aspect of trust behavior (as defined above). That is, given that the decision to trust is inherently associated with uncertainty about another's trustworthiness and the possibility of a loss (resulting from betrayal), an individual's general attitudes toward uncertainty (i.e., risk aversion) and potential losses (i.e., loss aversion) should affect the willingness to trust another. From a personality perspective, this implies that traits driving risk and loss aversion, namely anxiety-related traits, might likewise account for individual variation in trust behavior.

Furthermore, as follows from the uncertainty aspect of trust, trustors can never be completely sure about the trustee's trustworthiness, but have to estimate the probability of trust appreciation. These *trustworthiness expectations* (II; commonly called *beliefs* in

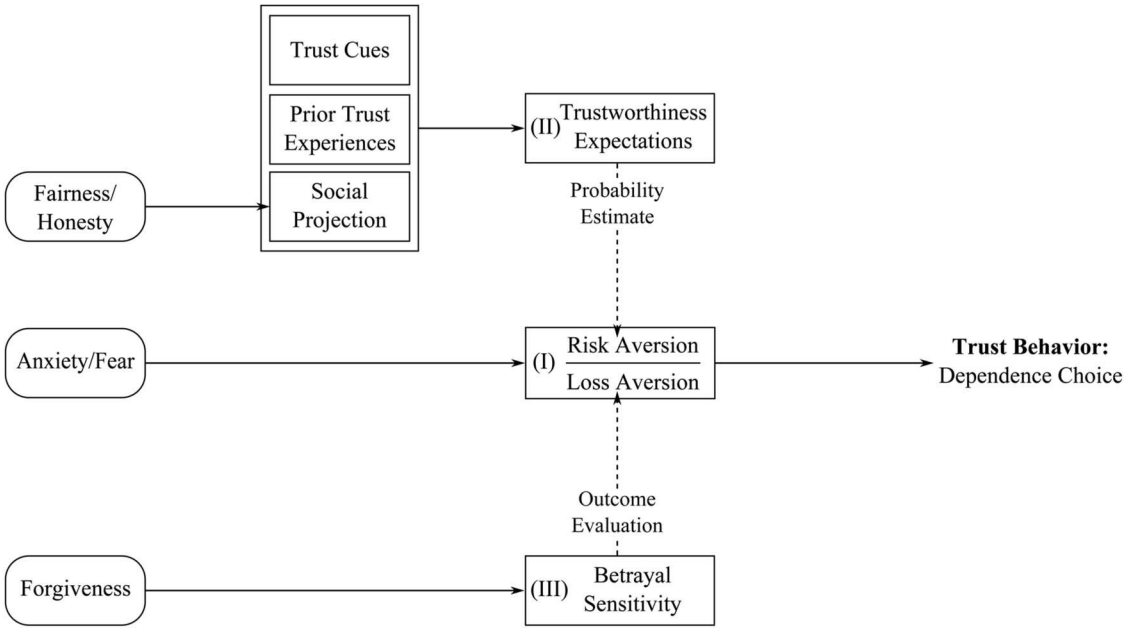

Figure 1. Graphical illustration of the three core components of trust behavior as distilled from the literature and their underlying determinants. Solid arrows reflect causal relationships; dashed arrows reflect information input. Personality traits are framed with round boxes.

economics) might be inferred based on different sources of information (depending on their availability): trust cues, prior trust experiences, and social projection. Trust cues are pieces of evidence available in the environment. They can either refer to the situation at hand (e.g., the temptation to betray) or to the trustee's person (e.g., her appearance). By contrast, prior trust experiences and social projection refer to internal sources of information specific to the trustor. More specifically, prior trust experiences mirror an individual's learning history in similar situations; social projection implies that a trustor's own trustworthiness (in terms of her trait fairness and honesty) might be a source of trustworthiness expectations—thus implying another potential source of individual variation in trust behavior besides risk and loss aversion.

Finally, as follows from the risk aspect of trust, trustors have to accept the possibility of a loss due to another person's betrayal. The way an individual perceives the severity of a loss resulting from betrayal (compared to a loss resulting from nature) denotes her *betrayal sensitivity* (III) which should, in turn, influence her general willingness to trust. In terms of a more basic trait, betrayal sensitivity arguably relates to trait forgiveness, thus representing an additional source accounting for individual differences in trust behavior.

Integrating the three components sketched so far, we propose that trustworthiness expectations provide the (subjective) probability input on which an individual's risk aversion operates whereas betrayal sensitivity provides the (subjective) outcome input on which an individual's loss aversion operates (cf. Figure 1). In what follows, we will describe these three determinants of trust in detail and elaborate on the underlying rational of the current framework.

I. Attitudes Toward Risky Prospects: Risk Aversion and Loss Aversion

Trust behavior as risky choice. As defined above, trust behavior refers to a risky choice of depending on another versus maintaining control over a personal resource. Correspondingly, trust behavior has frequently been described in terms of risk-taking versus risk-avoidance (e.g., Bohnet & Zeckhauser, 2004; Boon & Holmes, 1991; Das & Teng, 2004; Evans & Krueger, 2011; Ross & LaCroix, 1996; Rousseau et al., 1998; Ullmann-Margalit, 2004). As such, the decision to trust seems to be conceptually similar to a more general decision under risk which is characterized by four basic ingredients (in the decision making literature): the probability of a gain (positive outcome), the value or utility of the gain, the probability of a loss (negative outcome), and the value or utility of the loss (Payne, 1973; Slovic & Lichtenstein, 1968). In a situation of trust, the probabilities of gain versus loss reflect the probabilities of trust appreciation versus trust betrayal—that is, the trustee's trustworthiness. In case of a trustworthy interaction partner, the probability of a gain is higher (and the probability of a loss lower) than in case of an untrustworthy interaction partner. However, as stated above, the probabilities of gain versus loss are inherently unknown and thus have to be estimated or somehow inferred.

In terms of potential outcomes (of trust vs. distrust), the amount to gain reflects the positive utility resulting from trust appreciation whereas the amount to lose reflects the negative utility resulting from trust betrayal (cf. Evans & Krueger, 2011)—with the latter essentially reflecting the vulnerability aspect of trust behavior. Note, however, that recent evidence on the expressive nature of trust behavior (i.e., consideration of immediate rewards from the trusting act itself; Dunning et al., 2014, 2012; Dunning & Fetchenhauer, 2010, 2011) suggests that the amount to gain might also refer to the positive feelings associated with trusting and the communication of respect for another's moral character. Referring to the babysitter example from above, the gain hence corresponds to the positive feelings related to trusting the babysitter and spending an evening out with one's friends; the loss, in turn, corresponds to the potential "bad things" that might happen in the case the babysitter does not take good care of the child. In general, trustors are thus "confronted with an ambiguous path, a path that can lead to an event perceived to be beneficial [. . .] or to an event perceived to be harmful" (Deutsch, 1962, p. 303).

Supporting this idea that the decision to trust is conceptually similar to a decision under risk, evidence shows that individuals condition their trust behavior on both the probabilities of and the outcomes related to trust appreciation versus trust betrayal. That is, on the one hand, individuals indicated a higher willingness to trust a stranger when the chance of encountering a trustworthy interaction partner was 80% as compared to when the chance was only 46% (Fetchenhauer & Dunning, 2012)—that is, when it was "not too risky" to trust (Courtois & Tazdaït, 2012, p. 377). On the other hand, individuals were more likely to trust when trusting was associated with a high (as compared to a low) potential gain, but less likely to trust when trusting was associated with a high (as compared to a low) potential loss (Dufwenberg & Gneezy, 2000; Evans & Krueger, 2011, 2014; Goto, 1996; Lenton & Mosley, 2011; Malhotra, 2004; Snijders & Keren, 1999, 2001)—a behavioral tendency that is already apparent in children and adolescents (van den Bos, Westenberg, van Dijk, & Crone, 2010). From a dispositional point of view—and in line with the most widely accepted framework model of risky choice, prospect theory (Kahneman & Tversky, 1979, 1984)—the risk-taking aspect of trust behavior might hence depend on two determinants: (a) an individual's *risk aversion* (i.e., the willingness to take a risk as a function of the probabilities of gain vs. loss) and (b) an individual's *loss aversion* (i.e., the willingness to take a risk as a function of the relation between positive and negative outcomes). In the following, we will describe both components in detail.

Risk aversion. Risk aversion defines "an individual's dispositional tendency to evaluate a prospect with (positive) probabilistic outcomes as having a value lower than (i.e., risk aversion), equal to (i.e., risk neutrality), or higher than (i.e., risk seeking) its expected value" (Glöckner & Hilbig, 2012, p. 547). In other words, risk aversion captures an individual's preference for choosing a sure (positive) outcome over a potentially higher, but risky outcome. For example, given the choice between a safe gain of $10 and winning $20 with a probability of 50% (otherwise nothing), a risk-averse individual should prefer the safe gain whereas a risk-seeking individual should prefer the risky gamble. Transferred to the decision to trust, a risk-averse individual should hence require a larger subjective probability that the other is trustworthy (and therefore likely to honor trust) before actually trusting. In turn, a risk-seeking individual should be willing to trust even if she expects a smaller probability that the other is trustworthy (and therefore unlikely to honor trust). Following from this reasoning, dispositional risk aversion should be a determinant of trust behavior.

Supporting the influence of dispositional risk aversion on trust behavior, several studies identified risk aversion as a significant predictor of investments in the trust game (Altmann, Dohmen, & Wibral, 2008; Bigoni et al., 2013; Fetchenhauer & Dunning, 2012; Karlan, 2005; Lönnqvist, Verkasalo, Walkowitz, & Wichardt, 2010; Sapienza et al., 2013; Schechter, 2007). Specifically, individuals willing to entrust large amounts of their endowment (i.e., trusting individuals) were also more willing to take a risk in a lottery or in a financial decision than individuals willing to entrust only small amounts or nothing of their endowment (i.e., distrusting individuals).[4] Correspondingly, the term "risk" has been found to be most frequently associated with trust behavior in the trust game (Dunning et al., 2012). However, in contrast to these findings, the effect of risk aversion on trust behavior could not be consistently shown in previous research (Ashraf, Bohnet, & Piankov, 2006; Ben-Ner & Halldorsson, 2010; Corcos et al., 2012; Dunning et al., 2014; Eckel & Wilson, 2004; Etang, Fielding, & Knowles, 2011; Houser, Schunk, & Winter, 2010; Macko, Malawski, & Tyszka, 2014).

Strikingly though, researchers still disagree on how dispositional risk aversion can be appropriately assessed (and conceptualized). That is, several different measures of risk aversion have been proposed, including diverse self-report questionnaires (e.g., Dohmen, Falk, Huffman, Sunde, et al., 2011; Meertens & Lion, 2008; E. U. Weber, Blais, & Betz, 2002; Zuckerman, Eysenck, & Eysenck, 1978) as well as various behavioral measures such as the multiple price list method (Holt & Laury, 2002), the Balloon Analogue Risk Task (BART; Lejuez et al., 2002), the card gamble task (Eckel & Wilson, 2004), and the Bomb Risk Elicitation Task (BRET; Crosetto & Filippin, 2013). Given that (at least some of) these measures are only weakly related to each other (e.g., Ben-Ner & Halldorsson, 2010; Eckel & Wilson, 2004; Lönnqvist et al., 2010), it is questionable whether they actually measure the same construct (i.e., risk aversion).

In any case, besides measurement-related issues, research points to a high domain-specificity of risk aversion (i.e., financial, health/safety, recreational, ethical, and social; Blais & Weber, 2006; E. U. Weber et al., 2002), thus shedding doubts on whether risk aversion is a stable tendency across different contexts. In turn, it is questionable whether nonsocial risk aversion (e.g., lottery-based, financial) should predict trust-related risk-taking at all. Overall, trust researchers should bear these issues related to the assessment of risk aversion (i.e., heterogeneity of measures and domain-specificity) in mind and carefully select measures that are informative and appropriate regarding the specific research question under scrutiny. In the end, even the inconsistent findings on the relation between risk aversion and trust behavior do not necessarily contradict the idea that trust behavior is (to some extent) an expression of a dispositional risk-taking tendency (cf. Table 1). Nonetheless, in light of current knowledge, it seems unlikely that trust decisions are invariably influenced by dispositional risk aversion. We will later elaborate on how the interplay of the different trust determinants might account for a weak influence of risk aversion on trust under certain circumstances.

Loss aversion. In addition to risk aversion, *loss aversion* denotes another important aspect of risky choice, referring to the utilities of the potential outcomes rather than to their probabilities. That is, whereas risk aversion describes the willingness to take a risk as a function of the probabilities of gain versus loss, loss aversion describes the willingness to take a risk as a function of the relation between a gain and a loss (i.e., the outcomes). According to prospect theory (Kahneman & Tversky, 1979, 1984), loss aversion captures the intuition that a loss of X is more aversive than a gain of X is attractive. In terms of an individual difference variable, loss aversion can thus be defined as "the propensity for losses to loom larger than gains [of comparable magnitude]" (Bibby & Ferguson, 2011, p. 263; see also Gächter, Johnson, & Herrmann, 2010). Correspondingly, an individual with high loss aversion is assumed to evaluate a loss of X to be more aversive than a gain of X attractive—thus placing much more weight on a potential loss compared to a potential gain of equal magnitude (cf. Shelley, 1994). In contrast, an individual with low loss aversion is assumed to evaluate a loss of X as rather equally aversive as a gain of X attractive—thus placing comparable weights on both potential loss and potential gain.

Regarding the decision to trust, loss aversion might influence how an individual evaluates the potential gain resulting from honored trust (and the trusting act itself) in relation to the potential loss resulting from betrayed trust. On the one hand, trusting is typically accompanied by a potential gain in the case the trustee behaves trustworthily, thus implying that the trustor benefits from trusting in case the trustee honors the trust. In addition, the positive feelings associated with trusting might constitute a gain for the individual, regardless of the trustee's behavior. On the other hand, trusting as defined above is also necessarily accompanied by a potential loss resulting from the trustee's opportunity to betray, thus implying that the trustor suffers from trusting in case the trustee betrays. Referring to the babysitter example from above, one might spend an enjoyable evening out with friends if the babysitter takes good care of the child (i.e., behaves trustworthily), thus gaining from the trust placed in the babysitter beyond the mere positive feelings associated with the trusting act. At the same time, one also faces the risk of experiencing a loss (i.e., something bad happening to the child) in case the babysitter does not take her job seriously (i.e., behaves untrustworthily). Depending on the relation between the utilities of potential gain versus potential loss of trusting—in our example, the positive utility of trusting the babysitter and spending an evening out with friends versus the negative utility of whatever might occur to the child—individuals should hence be more or less willing to trust. In line with this idea, previous research already considered loss aversion as an underlying determinant of trust behavior (Aimone & Houser, 2012; Bohnet, Herrmann, & Zeckhauser, 2010; Bohnet & Meier, 2005). However, empirical evidence on the link between dispositional loss aversion and trust is still missing, thus remaining a quest for future research. Hence, it seems advisable for future research to consider both attitudes toward risky prospects (i.e., risk aversion and loss aversion) as distinct (risk-related) aspects underlying trust decisions (cf. Table 1).

[4] Note that some authors interpreted this effect of risk aversion on trust behavior in terms of a limitation of the trust game to appropriately disentangle trust from risk (Karlan, 2005; Sapienza et al., 2013; Schechter, 2007). However, in line with previously voiced arguments, we maintain that this effect demonstrates the (theoretically reasonable) risk-taking aspect of trust behavior (cf. Altmann et al., 2008; Dohmen, Falk, Huffman, & Sunde, 2011; Fehr, 2009; Lönnqvist et al., 2010).

Personality traits linked to risk and loss aversion. Linking risk aversion and loss aversion to general personality traits, both have most prominently been associated with trait anxiety and fear[5] (cf. Figure 1). With regard to risk aversion, evidence shows associations with both dispositional anxiety (Butler & Mathews, 1987; Lorian & Grisham, 2010; Maner & Schmidt, 2006; Maner et al., 2007; Wray & Stone, 2005) and dispositional fear (Lerner & Keltner, 2001; Maner & Gerend, 2007) as well as with anxiety- and fear-related basic traits (Glöckner & Hilbig, 2012; Lauriola & Levin, 2001; Lee, Ogunfowora, & Ashton, 2005; Weller & Thulin, 2012; Weller & Tikir, 2011) and pathological anxiety (Giorgetta et al., 2012; Lorian & Grisham, 2010, 2011; Lorian, Mahoney, & Grisham, 2012). Unlike in the case of risk aversion, evidence on the connection between trait anxiety/fear and loss aversion is rather scarce. Nonetheless, like risk aversion, loss aversion has also been linked to anxiety-related traits (Bibby & Ferguson, 2011) as well as to fear (Camerer, 2005; McCarter, Rockmann, & Northcraft, 2010). Overall, these links imply that dispositional anxiety might account for individual variation in trust behavior (through its link to risk and loss aversion). However, given the lack of evidence on the relation between dispositional anxiety and trust behavior, further research is needed to clarify whether findings on the relation between dispositional anxiety and risk/loss aversion can actually be generalized to the trust context.

Preliminary summary of (I) attitudes toward risky prospects. Based on the reasoning that the decision to trust reflects a decision under risk, the present framework includes two attitudes toward risky prospects as underlying determinants of an individual's willingness to trust, distilled from the extant literature: risk aversion and loss aversion (cf. Figure 1). Whereas risk aversion denotes the willingness to take a risk as a function of the probabilities of gain versus loss, loss aversion denotes the willingness to take a risk as a function of the relation between the utilities of the potential gain versus the potential loss. Thus, each attitude refers to a specific aspect of the risky choice of trusting (i.e., probabilities and outcomes). Depending on the importance individuals assign to either aspect in the specific trust situation, each might drive the decision to trust more or less strongly. In terms of an underlying personality characteristic, both are linked to trait anxiety/fear.

Admittedly, so far we presumed that individuals are familiar with the probabilities of trust appreciation versus trust betrayal—since we merely referred to risk aversion as an individual's tendency to take a risk as a function of (known) probabilities of gain versus loss. However, as reasoned above, this is typically not the case because of the inherent uncertainty about the trustee's trustworthiness. Individuals thus have to estimate these probabilities based on the information available in the situation. Moreover—given that the potential loss of trusting results from another person's betrayal (rather than from bad luck)—trusting might not only involve a material loss, but also a psychological loss because of the manner in which the loss comes about (i.e., an action of another individual vs. chance). In consequence, the evaluation of the potential outcomes (especially losses) might not be as straightforward as in a nonsocial decision under risk such as gambling. Based on these notions, we propose *trustworthiness expectations* and *betrayal sensitivity* as two additional determinants of trust behavior (component II and III, respectively, in Figure 1). Whereas trustworthiness expectations essentially reflect individuals' probability estimates of gain versus loss, betrayal sensitivity should underlie individuals' evaluation of the potential outcomes (particularly losses). Both determinants will be described in the following.

II. Trustworthiness Expectations

Following from the inevitable uncertainty about the probabilities of trust appreciation (gain) versus trust betrayal (loss), a trustor has to infer another's trustworthiness before deciding whether it is reasonable to trust or not (e.g., Anh, Pereira, & Santos, 2011; Bacharach & Gambetta, 2001; Chang, Doll, van't Wout, Frank, & Sanfey, 2010; Eckel & Wilson, 2004; Hill & O'Hara, 2006; Kramer, 2010; Macy & Skvoretz, 1998; Snijders & Keren, 1999, 2001; Yamagishi & Yamagishi, 1994). Depending on the specific trust situation, these inferences might, for example, refer to whether the trustee provides trustworthy information, behaves cooperatively in a negotiation or an exchange of resources, or refrains from (emotionally) harming the trustor.[6] In any case, in interactions with strangers, a trustor must draw these inferences in the absence of reliable information about the interaction partner's trustworthiness. To nevertheless form an expectation about the other's likely behavior, the trustor can consider different (more or less implicit) sources of information, namely (a) trust cues, (b) prior trust experiences, and (c) social projection.

Trust cues. *Trust cues* are in some way observable or given pieces of evidence a trustor might use to draw inferences about a trustee's trustworthiness in a specific situation. In terms of Brunswik's (1952) lens model, trust cues are observable cues available in the trustor's environment, yielding probabilistic information about an interaction partner's likely trustworthiness. According to a person-situation-interaction perspective (e.g., Funder, 2008), one can differentiate between two types of trust cues: personal and situational. Whereas *personal trust cues* refer to individual characteristics of the trustee (i.e., outward appearance, reputation, or social category), *situational trust cues* refer to features of the specific trust situation (i.e., temptation to betray or availability of potential sanctions).

First off, personal trust cues denote characteristics of the trustee predicting her trustworthiness or cooperativeness, respectively. In interactions with strangers, personal trust cues mainly refer to the trustee's outward appearance, including facial features (Stewart et al., 2012; Todorov, Baron, & Oosterhof, 2008; Todorov, Pakrashi, & Oosterhof, 2009), facial expressions (Campellone & Kring, 2013; Eckel & Wilson, 2003; Krumhuber et al., 2007; Little, Jones, DeBruine, & Dunbar, 2013; Scharlemann, Eckel, Kacelnik, & Wilson, 2001; Shinada et al., 2010; Sofer, Dotsch, Wigboldus, & Todorov, 2015; Yu, Saleem, & Gonzalez, 2014), facial expressiv-

[5] Specifically, anxiety refers to the reaction of approaching to danger whereas fear refers to the reaction of escaping from danger (McNaughton, 2011).

[6] The diverse nature of trustworthiness expectations (determined by the specific trust situation) implies that these inferences might refer to both *epistemic trust* concerns (expectations that an informant provides reliable/correct information; e.g., P. L. Harris, 2007; Koenig & Harris, 2005; Landrum et al., 2015; Shafto et al., 2012) and *emotional trust* concerns (expectations that another refrains from causing emotional harm to the trustor; e.g., Betts, Rotenberg, & Trueman, 2009; Corriveau & Harris, 2010; Johnson-George & Swap, 1982; Rotenberg, 2010) as typically distinguished in developmental psychology.

ity (Boone & Buck, 2003), and body language (DeSteno et al., 2012; Naumann, Vazire, Rentfrow, & Gosling, 2009). For example, individuals from different cultures seem to agree that faces with high inner eyebrows, pronounced cheekbones, wide chins, and shallow nose sellion appear more trustworthy than faces with low inner eyebrows, shallow cheekbones, thin chins, and deep nose sellion (Birkás, Dzhelyova, Lábadi, Bereczkei, & Perrett, 2014; Todorov et al., 2008). Moreover, individuals evaluated facial expressions (e.g., smiling) and body language cues (e.g., energetic and tense stance) as indicating (high) Agreeableness (Naumann et al., 2009)—a potential predictor of high cooperativeness (Denissen & Penke, 2008). Overall, individuals seem to automatically use appearance-based personal trust cues to predict the trustworthiness of unknown interaction partners (Bonnefon, Hopfensitz, & De Neys, 2013; De Neys, Hopfensitz, & Bonnefon, 2015) and, in turn, to condition their trusting behavior on these judgments (Chang et al., 2010; DeSteno et al., 2012; Oda, Naganawa, Yamauchi, Yamagata, & Matsumoto-Oda, 2009; Posten, Ockenfels, & Mussweiler, 2014; van 't Wout & Sanfey, 2008).

Note, however, that reliance on trust cues has no immediate implication for whether or not the trustworthiness judgment ultimately formed is, in fact, correct or accurate (cf. Table 1). Indeed, whereas some studies suggest that people are somewhat accurate in predicting the trustworthiness of others on the basis of facial cues (D. S. Berry, 1990; Bond, Berry, & Omar, 1994; Little et al., 2013; Stirrat & Perrett, 2010), other studies indicate the opposite (e.g., Bonnefon et al., 2013; Efferson & Vogt, 2013; Manson, Gervais, & Kline, 2013; Rule, Krendl, Ivcevic, & Ambady, 2013). In turn, the accuracy of predictions does not seem to be moderated by individuals' Theory of Mind, that is, the ability to infer others' mental states (Sylwester, Lyons, Buchanan, Nettle, & Roberts, 2012). Nevertheless, "trustworthiness judgments from faces reflect inferences of behavioral intentions that signal approach/avoidance behaviors" (Todorov, 2008, p. 220).

Besides outward appearance cues, "reputation provides a good opportunity for trust to prosper" (Yamagishi & Yamagishi, 1994, p. 138; see also Klapwijk & Van Lange, 2009; Manapat, Nowak, & Rand, 2013). Whenever reputational information is available, a trustor might hence consider the trustee's reputation to form a trustworthiness judgment. Correspondingly, evidence suggests that individuals actually condition their trust behavior on reputational information. Specifically, individuals are more willing to trust people with a positive reputation than those with a negative reputation (Albert, Güth, Kirchler, & Maciejovsky, 2007; Barclay, 2004; Boero, Bravo, Castellani, & Squazzoni, 2009; Bohnet & Huck, 2004; Bracht & Feltovich, 2009; Charness, Du, & Yang, 2011; Delgado, Frank, & Phelps, 2005; Fehrler & Przepiorka, 2013; Gambetta & Przepiorka, 2014; Keser, 2003; King-Casas et al., 2005; Manapat & Rand, 2012; Masuda & Nakamura, 2012; Rezlescu, Duchaine, Olivola, & Chater, 2012). Hence, individuals seem to use reputational information as a vital cue in their decision to trust.

Regarding trust among strangers, the role of reputational information seems to be particularly relevant in the context of online purchase (e.g., Kollock, 1999; Resnick & Zeckhauser, 2002). For example, online purchase systems (such as eBay or Amazon) specifically implement reputation systems by collecting buyers' and sellers' ratings of each other after each transaction and, in turn, making them available to all users. Indicating the effectiveness of such reputation systems, trustors seem to place higher trust in favorably rated sellers than in unfavorably rated sellers, indicated by a larger probability of sale and higher average selling prices for sellers with a positive reputation (Przepiorka, 2013). However, information about a seller's reputation does not seem to suppress the importance of outward appearance on trust behavior when both trust cues are presented in combination (Bente, Baptist, & Leuschner, 2012).

Finally, social category represents another personal trust cue relevant for the occurrence of trust among strangers. For example, trustors place higher trust in strangers who are in-group members (due to expecting altruistic and fair treatment)—a phenomenon termed *group-based trust* (Foddy, Platow, & Yamagishi, 2009; Platow, Foddy, Yamagishi, Lim, & Chow, 2012). In particular, meta-analytic evidence indicates that individuals have more optimistic expectations regarding in-group members' trustworthiness (Balliet, Wu, & De Dreu, 2014)—thus rendering high trust reasonable. Similarly, knowledge about another's social category can trigger social stereotypes which, in turn, may be associated with specific trustworthiness expectations (Brewer, 2008). In general, the tendency to predict another's trustworthiness based on social category information is already apparent in children (Landrum, Eaves, & Shafto, 2015). Altogether, a trustee's social category is hence another potential source of information people might use to infer the trustworthiness of an unknown interaction partner.

In addition to personal trust cues, trustors might also consider situational trust cues in their probability estimates of trust appreciation versus trust betrayal. Unlike personal trust cues, situational trust cues refer to characteristics of the situation that might affect an interaction partner's trustworthiness independent of (or in some interaction with) her general trustworthiness. In interdependence situations—as the one of trust—several characteristics of the situation can be distinguished (e.g., dependence and power imbalances, conflicts of interest, and behavioral control; Kelley et al., 2003). Most prominently, research on trust considered a trustee's temptation to betray, that is, the difference in the trustee's payoff between honoring and betrayal of trust (cf. Evans & Krueger, 2011), as a vital situational trust cue. Specifically, the temptation to betray refers to the trustee's conflict of interest (Kelley et al., 2003). Thus, it should be relevant to a trustor's expectation about the trustee's trustworthiness because it reasonably affects the trustee's motivation to betray rather than to honor trust (Balliet & Van Lange, 2012; Evans, Athenstaedt, & Krueger, 2013; Evans & Krueger, 2011, 2014; James, 2002; Murray & Holmes, 2009; Snijders & Keren, 1999, 2001; Yamagishi, Kanazawa, Mashima, & Terai, 2005). In situations in which trust betrayal leads to a considerably higher outcome for the trustee than trust appreciation (i.e., high temptation to betray), a trustor might expect a relatively high probability of trust betrayal and should, in turn, be less willing to trust the other. By contrast, in situations in which trust betrayal only leads to a slightly higher outcome for the trustee than trust appreciation (i.e., low temptation to betray), a trustor might expect a relatively low probability of trust betrayal and should, in turn, be more willing to trust the other. In line with this reasoning, empirical findings demonstrate a decline in trust behavior when the trustee's temptation to betray increases (Evans & Krueger, 2011, 2014; Snijders & Keren, 1999, 2001). However, trustors appear to underestimate the impact of the temptation to betray on the trustee's trustworthiness. Specifically, focusing primarily on their own potential outcomes, trustors consider a trustee's temptation only when the

potential losses and gains associated with the trusting choice are favorable for themselves (Evans & Krueger, 2011, 2014).

Similar to the temptation to betray, the presence versus absence of potential sanctions for betrayal constitutes another situational trust cue. Clearly, potential sanctions in case of betrayal decrease the attractiveness of betrayal for the trustee, thus allowing the trustor to expect a higher probability of trust appreciation (Bacharach & Gambetta, 2001; Yamagishi, 1986, 1988; see also Balliet, Mulder, & Van Lange, 2011, for a meta-analytic review on the cooperation-promoting effect of potential punishment). Consistent with this idea, the presence of potential sanctions for trustees' betrayal (e.g., payoff reduction or negative reputation) increased trustors' optimism about the trustees' trustworthiness and, in consequence, their willingness to trust (Bigoni et al., 2013; Bohnet & Baytelman, 2007; Charness, Cobo-Reyes, & Jimenez, 2008; Matthews, Kordonski, & Shimoff, 1983; Stiff, 2008; Vollan, 2011). Hence, it is reasonable to consider potential sanctions (or their absence) as another situational trust cue, beyond the temptation to betray.[7]

Prior trust experiences. Following from the notion that trust refers to a basic social behavior, it is reasonable to expect that people can fall back on a substantial number of experiences related to trust behavior. Such *prior trust experiences* should, in consequence, influence people's trust behavior in comparable situations (Bicchieri, Duffy, & Tolle, 2004; Blair & Stout, 2001; Bohnet & Huck, 2004; Deutsch, 1962; Glanville & Paxton, 2007; Rotter, 1967, 1980; Snijders & Keren, 1999; Tullberg, 2008; Van Lange, Vinkhuyzen, & Posthuma, 2014; Yu et al., 2014), most arguably through affecting their expectations about others' trustworthiness. Especially in situations in which trust cues are absent or ambiguous, prior trust experiences (collected in similar situations) might guide one's trust behavior in terms of an *a priori trustworthiness expectation*. For example, an individual who had positive experiences with buying a car second-hand from an unknown seller (i.e., encountering a cooperative interaction and ending up with a car of high quality) might be willing to do so again in the future—even from another seller, thus expecting this other (unknown) seller to be similarly trustworthy.

Supporting this reasoning, buyers whose trust had been betrayed in an online transaction generalized their reduced trust to all sellers whereas buyers whose trust had been rewarded maintained their general trust in the sellers (Bolton, Katok, & Ockenfels, 2004). Hence, trusting is arguably a "learned behavior" (Blair & Stout, 2001, p. 1742; see also Van Lange, 2015) to some extent as it will be based on previous experiences with others' trustworthiness (Courtois & Tazdaït, 2012; Fang, Kimbrough, Valluri, Zheng, & Pace, 2002; Landrum et al., 2015; Rothmund, Gollwitzer, Bender, & Klimmt, 2015; van den Bos, van Dijk, & Crone, 2012). Note that the accuracy of trustworthiness judgments based on prior trust experiences will essentially depend on whether or not the generalization (from prior experiences to the current situation) is justified. Typically, this will be more often the case if the current trustee is somehow similar to the one(s) encountered previously. In general, future research might profit from taking into account the potential influence of prior trust experiences on trust behavior (cf. Bellemare & Kroger, 2007) and, in turn, consider whether individuals have participated in previous studies using similar paradigms (e.g., the trust game) in which they have received feedback on another's trustworthiness (cf. Table 1).

Social projection. Besides trust cues and prior trust experiences, a third (trait-based) source of information people might consider in their probability estimates of others' trustworthiness is *social projection*. Social projection is a "judgmental heuristic that leads people to expect that others will behave as they themselves do" (Krueger & Acevedo, 2005, p. 18). As recently discussed by Krueger and colleagues (Krueger & Acevedo, 2005; Krueger, DiDonato, & Freestone, 2012; Krueger, Massey, & DiDonato, 2008; Krueger, 2007), social projection represents a reasonable antecedent of trust.[8] According to this view, people predict another's trustworthiness (or cooperativeness, respectively) by projecting their own cooperative preferences onto the other. Based on this prediction of another's trustworthiness, people might then decide whether it is reasonable to trust. Correspondingly, a cooperative individual should expect others to be cooperative (i.e., trustworthy) as well, thus considering trust reasonable due to a high expected probability of trust appreciation. In contrast, an uncooperative individual should expect others to be uncooperative as well, thus considering trust unreasonable due to a high expected probability of betrayal.

In line with the idea that one's own cooperativeness influences trust behavior (i.e., through social projection), different variants of other-regarding preferences such as altruism, unconditional kindness, fairness, and inequality aversion have repeatedly been considered relevant predictors of trust (Ashraf et al., 2006; Cox, 2004; Derks, Lee, & Krabbendam, 2014; Dunning et al., 2012; Fehr, 2009; Hong & Bohnet, 2007; Johansson-Stenman, Mahmud, & Martinsson, 2013; Krueger et al., 2008; Lehmann-Waffenschmidt & Leipold, 2011; Mansbridge, 1999; Sapienza et al., 2013). Indeed, evidence shows that individuals with a prosocial social value orientation (SVO) or a high willingness to cooperate in the dictator or the prisoner's dilemma game were more likely to trust than individuals with a proself SVO or a low willingness to cooperate in either game (Ashraf et al., 2006; Ben-Ner & Halldorsson, 2010; Bohnet & Baytelman, 2007; Chaudhuri & Gangadharan, 2007; Cox, 2004; Etang et al., 2011; Kanagaretnam, Mestelman, Nainar, & Shehata, 2009; Macko et al., 2014; Sapienza et al., 2013; Yamagishi et al., 2013, 2012). In the same vein, findings from the trust game demonstrate that high trustworthiness is typically accompanied by a high willingness to trust. That is, individuals who honored (rather than betrayed) trust as a trustee also en-

[7] Note, however, that potential sanctions should only increase trust concerning others' external motivation to honor trust, but might even decrease trust concerning others' intrinsic motivation to do so (Chen, Pillutla, & Yao, 2009; Mulder, van Dijk, De Cremer, & Wilke, 2006). Besides, the effectiveness of potential sanctions to increase trust might differ across societies as a function of generalized trust in others, that is, trust propensity (Balliet & Van Lange, 2013).

[8] Note that social projection has also been (implicitly) considered in previous trust theories. For example, Sapienza et al. (2013) propose that belief-based trust (in the sense of trustworthiness expectations) is strongly affected by a trustor's own behavior. Similarly, Hill and O'Hara (2006) assume that people facing a trust decision imagine others to behave as they themselves would behave. Finally, Brewer (2008) refers to social projection as a mediator of intragroup trust. Likewise, social projection has been discussed early on as a determinant of expectations about a partner's willingness to cooperate in the prisoner's dilemma game (Dawes, McTavish, & Shaklee, 1977; Kelley & Stahelski, 1970; Orbell & Dawes, 1991). Yet, in this context, trustworthiness expectations have only been assessed a posteriori to actual trusting choices, thus considering social projection as a postchoice process rather than as a (prechoice) determinant of expectations and corresponding behavior.

trusted relatively large amounts of money as a trustor (Altmann et al., 2008; Chaudhuri & Gangadharan, 2007; De Neys et al., 2015; Evans & Revelle, 2008; Schechter, 2007; Vyrastekova & Garikipati, 2005; Yamagishi et al., 2013).

However, given that the investment in the trust game affects another person's (i.e., the trustee's) outcome, it does not represent a pure measure of trust, but also incorporates aspects of cooperation (Ben-Ner & Putterman, 2009; Burks, Carpenter, & Verhoogen, 2003; Cox, 2004; Evans et al., 2013; Holm & Nystedt, 2008; Kamas & Preston, 2012; McEvily, Radzevick, & Weber, 2012; Vollan, 2011; Vyrastekova & Garikipati, 2005; Yamagishi et al., 2013). Hence, evidence linking other-regarding preferences to trust behavior in the trust game cannot conclusively corroborate a mechanism of social projection, but might reflect the cooperation-related aspect of trust game investments. Ruling out this alternative explanation, recent evidence based on the faith game—a game sought to provide a *pure* measure of trust in another's cooperativeness (Kiyonari, Yamagishi, Cook, & Cheshire, 2006; Kiyonari & Yamagishi, 1999)[9]—also pointed to a significant relation between individuals' own cooperativeness (in different economic games) and trust behavior (Yamagishi et al., 2013). Similarly, trait cooperativeness has been specifically identified as a predictor of trustworthiness expectations (Thielmann & Hilbig, 2014). That is, compared to uncooperative individuals, cooperative individuals were more optimistic about an unknown other's fairness and, in turn, trustworthiness. In summary, various findings support the notion that social projection of one's own trustworthiness onto others might be a common source of information people consider in forming trustworthiness expectations. Note, however, that people will differ in their tendency to project their own characteristics onto others (Krueger & Acevedo, 2007). Correspondingly, people might also differ in their tendency to use social projection as a means of predicting others' trustworthiness.

Following from the idea that individuals' own trustworthiness affects their trustworthiness expectations (through social projection) and, in turn, their trust behavior, personality traits determining an individual's trustworthiness should predict trust behavior. In general, trustworthiness comprises two components, namely cooperation and honesty (cf. Van Lange, Van Vugt, Meertens, & Ruiter, 1998; Yamagishi & Yamagishi, 1994; Yamagishi, 1988). Most basically, cooperation is reducible to trait fairness (e.g., Arneson, 1982; Eek & Biel, 2003; Fehr & Schmidt, 1999; Güney & Newell, 2013; Kahneman, Knetsch, & Thaler, 1986; Rabin, 1993; Wilke, 1991), which represents an individual's willingness to share or give rather than to exploit others, even if there is no particular incentive to do so.[10] Honesty, in contrast, describes the degree to which one can rely on the word or promise of an individual (cf. López-Pérez, 2012). Correspondingly, people typically ascribe benign intentions to honest others (Ashton, Lee, & Son, 2000; Rosenberg & Sedlak, 1972; Van Lange & Kuhlman, 1994; Van Lange & Semin-Goossens, 1998; C. S. Wang, Galinsky, & Murnighan, 2009). Overall, both trait fairness and trait honesty are hence reasonable determinants of trustworthiness and—based on the idea of social projection—of individual differences in trust behavior (cf. Figure 1).

Note, however, that these two components should underlie trust behavior in somewhat different situations: Whereas fairness refers to trust behavior in situations in which the trustor depends on the trustee's cooperativeness (e.g., the willingness to return entrusted money in the trust game), honesty refers to trust behavior in situations in which the trustor depends on the trustee's word or promise (e.g., the willingness to truthfully report some relevant information). In any case, both components refer to an individual's trustworthiness and should therefore refer to a single underlying trait as determinant. Nevertheless, it seems advisable to consider which aspect of trustworthiness is more relevant for the trust situation at hand (fairness vs. honesty) and to choose corresponding methods in future studies concerned with the influence of social projection on trust decisions.

Preliminary summary of (II) trustworthiness expectations. As detailed above, we summarize three different sources of information individuals might use to predict an unknown other's trustworthiness (cf. Figure 1): (a) situational and personal trust cues available in a specific trust situation, (b) prior trust experiences collected in situations similar to the one at hand, thus reflecting an a priori trustworthiness expectation about unknown others in a specific situation, and (c) social projection of one's own trustworthiness, capturing the personality aspect underlying trustworthiness expectations in terms of trait fairness and trait honesty. Integrating these pieces of evidence, a trustor may arrive at some estimate of the probability of trust appreciation versus trust betrayal, thereby reducing uncertainty about the other's likely behavior. However, by definition, a trustor can never be completely certain about another's trustworthiness as long as the situation is one of trust rather than confidence (i.e., assurance without an involvement of uncertainty or risk, respectively; Luhmann, 1988), regardless of the number and the validity of the available trustworthiness-related pieces of evidence.

Concerning the accuracy of trustworthiness judgments, it appears that even children are somehow able to distinguish trustworthy from untrustworthy trustees based on different trust cues (see Mills, 2013, for a recent review). However, it should also be noted that (even in adults) the trustor's judgment does not necessarily

[9] In the faith game, individuals decide whether to receive an unknown amount of money a randomly matched dictator allocated to a recipient or to receive a fixed amount of money from the experimenter that is known to be less than half of the amount the dictator could divide. If an individual trusts in the fairness of the dictator, she should prefer the (potentially higher) amount allocated by the dictator over the fixed amount from the experimenter. But, if an individual distrusts in the fairness of the dictator, she should go for the fixed amount from the experimenter. Note that, in general, the trusting choice in the faith game has no effect on the joint or the partner's outcome.

[10] Besides fairness, previous research has considered altruism (i.e., personal sacrifice on behalf of others) and greed (i.e., high importance of own gains) as general traits underlying cooperation. However, because both altruism and greed have usually been defined in terms of dictator game giving—which basically reflects a measure of fairness—they are well captured in the concept of trait fairness (cf. Eek & Biel, 2003; Wilke, 1991). Similarly, scholars have proposed several specific traits as underlying factors of individual differences in cooperative behavior. Most prominently, cooperation has been linked to social value orientation, reflecting the weights people assign to their own and to others' outcomes in situations of interdependence (Messick & McClintock, 1968; for a meta-analytic overview see Balliet, Parks, & Joireman, 2009). Furthermore, inequality aversion (i.e., disliking differences between own and others' or own and average payoffs, respectively; Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) and strong reciprocity (i.e., high willingness to cooperate with cooperative others and to punish those who violate the norms of cooperation, at a personal cost; Gintis, Bowles, Boyd, & Fehr, 2003) have been discussed as specific traits underlying cooperation. However, all these concepts ultimately share a common core captured by what we mean by "fairness".

reflect the available evidence in terms of an optimal probability estimate. For example, individuals might consider only a subset of the available information due to time pressure, low motivation, or little need to reduce uncertainty about the other's trustworthiness.[11] They might use simplifying heuristics for probability estimation (Kahneman, Slovic, & Tversky, 1982) and/or different strategies for cue integration (Gigerenzer & Goldstein, 1996; Newell & Bröder, 2008), the accuracy of which will depend on the cue environment (Hogarth & Karelaia, 2007). Also, the probability estimate may be distorted by an individual's sensitivity to victimization (Gollwitzer, Rothmund, Alt, & Jekel, 2012; Gollwitzer, Rothmund, Pfeiffer, & Ensenbach, 2009; Gollwitzer, Rothmund, & Süssenbach, 2013), emotional state (J. R. Dunn & Schweitzer, 2005), and a *negativity bias* (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001), mirroring people's general tendency to overestimate the probability of negative events compared to positive or neutral ones (A. J. L. Harris, Corner, & Hahn, 2009). As the latter suggests, people may generally overestimate the probability of trust betrayal—which is in line with evidence indicating that people tend to generally underestimate others' trustworthiness (Fetchenhauer & Dunning, 2009, 2010; D. T. Miller, 1999). Unfortunately, though, there have been no attempts so far to integrate the literature on trustworthiness expectations and (cognitive processes of) judgments under uncertainty as typically addressed in the cognitive psychological literature. Thus, the question of how different sources of information are integrated and to what extent intuitive versus deliberate cognitive processes are at work (cf. Betsch & Glöckner, 2010; Glöckner & Witteman, 2010) is largely unanswered. Undoubtedly, the expected probability of another's (un)trustworthiness should strongly influence an individual's willingness to trust. To specify how exactly individuals form this probability estimate—and thus bridge the noteworthy gap between trust research and judgment research—remains a vital quest for future research (cf. Table 1). However, recent developments in (probabilistic) modeling of epistemic trust in children based on Bayesian inference (Shafto, Eaves, Navarro, & Perfors, 2012) seems to provide a valuable starting point and a fruitful (statistical) approach to close this gap.

III. Betrayal Sensitivity as a Determinant of Outcome Evaluation

In addition to the determinants introduced so far (i.e., (I) attitudes toward risky prospects and (II) trustworthiness expectations), we finally propose (III) *betrayal sensitivity* as a third component underlying trust behavior, particularly influencing individuals' evaluations of potential losses resulting from trust behavior. As discussed above, trusting is accompanied by risk (in the sense of a potential loss) due to the trustee's opportunity to betray and the trustor's inevitable uncertainty about the trustee's inclination to do so. This suggests that if trusting results in a loss, the loss can specifically be traced back to another person's betrayal (rather than to bad luck, for instance). A loss resulting from misplaced trust might hence go beyond a mere material loss, but also involve a psychological loss due to the way the loss came about (cf. Rabin, 1993). As Bohnet et al. (2010) reasoned based on prospect theory, "betrayal imposes an additional utility cost beyond monetary loss. That cost increases the more the likelihood of betrayal deviates from one's reference points of accustomed experience" (p. 812). In line with this reasoning, recent research suggests that people are, in general, less willing to take a risk when the risk results from human selfish and untrustworthy behavior (i.e., trust situation) as opposed to the risk resulting from bad luck (i.e., gamble situation)—a phenomenon called *betrayal aversion* (Aimone & Houser, 2011, 2012, 2013; Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Bohnet, Greig, Herrmann, & Zeckhauser, 2008; Bohnet & Zeckhauser, 2004; Corcos et al., 2012; Fehr, 2009; Hong & Bohnet, 2007; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005).

Supporting the idea of betrayal aversion, several studies have demonstrated that individuals distinguish between risks resulting from betrayal and risks resulting from bad luck. More specifically, one line of research showed that participants are less willing to take a risk in trust situations compared with nonsocial risk situations (Baumgartner et al., 2008; Bohnet et al., 2008; Bohnet & Zeckhauser, 2004; Kosfeld et al., 2005). In these studies, participants had to estimate the minimum acceptable probability of gaining the higher of two outcomes in a (trust or risk) game for which they would prefer the game over a sure (but relatively low) outcome. In the trust game, the final outcome was determined by another person; in the risk game, it was determined by nature. As expected, individuals indicated higher minimum acceptable probabilities in the trust game than in the risk game, thus indicating a lower willingness to take a risk in situations in which a poor outcome is produced by another's betrayal rather than the same outcome being due to bad luck. Of note, recent evidence on the biological basis of trust suggests that the positive effect of oxytocin on trust is due to a reduction in betrayal aversion (Baumgartner et al., 2008; Kosfeld et al., 2005) which, in turn, most likely results from a decrease in amygdala activation (e.g., Kirsch et al., 2005; see also Rilling & Sanfey, 2011, for an overview).

In addition to differences between trust and risk games, further evidence suggests that individuals are particularly motivated to prevent suffering from knowing that another person betrayed their trust (Aimone & Houser, 2012, 2013). That is, when participants were able to choose whether they want to know the decision of their assigned trustee or instead receive a payment according to a random draw from a separate pool of decisions identical to the pool of the trustee's decisions, they preferred to avoid knowing whether their trustee honored or betrayed their trust and rather chose to be paid according to the behavior of a randomly drawn trustee. Strikingly, if participants could not avoid knowing their partner's trustworthiness, they were less willing to trust in the first place. Irrational though this appears, people may thus be more willing to trust if they need not learn whether the trustee betrayed them. In real-life, however, one may argue that individuals cannot avoid receiving feedback about a trustee's trustworthiness (or betrayal, respectively), unless they mistrust in the first place (cf.

[11] Individuals usually differ in their motivation to reduce uncertainty as a function of their individual level of *ambiguity aversion* (i.e., the preference for known risks over unknown risks; Ellsberg, 1961). In a trust situation, individuals with high ambiguity aversion might hence gather as much information as possible to come up with a fairly reliable (and rather certain) prediction about an interaction partner's trustworthiness. Individuals with low ambiguity aversion, in contrast, might be satisfied with only gathering little information given that they should consider the state of uncertainty less aversive.

Fetchenhauer & Dunning, 2010). Nevertheless, the findings reviewed so far support the notion that individuals differentiate between different sources of risk (i.e., human selfishness vs. nature)—showing sensitivity toward betrayal (cf. Lehmann-Waffenschmidt & Leipold, 2011).[12]

Betrayal sensitivity. Based on these findings on betrayal aversion, *betrayal sensitivity*—an individual's tendency to attribute a greater severity to a loss resulting from human selfishness than to a formally equivalent loss resulting from nature—should be a vital determinant underlying trust behavior.[13] In particular, individuals might differ regarding their weighting of (objectively comparable) losses resulting from taking a social versus taking an asocial risk (cf. Table 1). That is, whereas some individuals might perceive the potential loss of trusting rather similar to the potential loss of gambling (i.e., low betrayal sensitivity), others might perceive the potential loss of trusting clearly more severe than a formally equivalent loss of gambling (i.e., high betrayal sensitivity). This implies another source of interindividual variation in trust decisions. In effect, betrayal sensitivity should be negatively related to trust because of its influence on the evaluation of outcomes, particularly losses.

Given that individuals high in betrayal sensitivity are assumed to ascribe a greater severity to losses resulting from nonappreciated trust than to comparable losses resulting from bad luck, betrayal sensitivity should arguably be driven by low interpersonal forgiveness (cf. Fehr, 2009). Forgiveness refers to the willingness to refrain from retaliatory actions when others are behaving unfairly. If one is quick to forgive, there is little need for betrayal aversion. Supporting this notion, betrayal has been found to induce reactions that are antithetical to forgiveness (i.e., grudge, vengeance, and retribution; Finkel, Rusbult, Kumashiro, & Hannon, 2002; Stouten, De Cremer, & van Dijk, 2006). Similarly, trust behavior following a transgression was predicted by trait forgiveness (Desmet, De Cremer, & van Dijk, 2011) as well as dispositional (dis)trust (Maltby et al., 2008; Walker & Gorsuch, 2002). We thus propose individual differences in betrayal sensitivity to be explainable by trait forgiveness. However, because betrayal sensitivity has not yet been operationalized as an individual difference construct, the relation between betrayal sensitivity and general personality traits (such as forgiveness) remains an open quest for future research.

In summary, we propose betrayal sensitivity as a third determinant underlying trust behavior (besides risk aversion/loss aversion and trustworthiness expectations; cf. Figure 1), especially affecting the evaluation of potential losses resulting from trust behavior. Individuals high in betrayal sensitivity are expected to be less willing to take a risk in a trust situation than in a comparable risky gamble situation (due to ascribing a higher negative utility to a loss resulting from betrayal than to a formally equivalent loss resulting from bad luck). Given that individuals high in betrayal sensitivity thus seem to particularly dislike nonappreciated trust, trait forgiveness should arguably drive individual differences in betrayal sensitivity—and thus trust.

Reward sensitivity. Note, however, that in addition to focusing on the potential losses associated with betrayed trust (as implied by the notion of betrayal sensitivity), some individuals may also place considerable attention on the potential reward inherent in a positive (i.e., trusting) social interaction. This corresponds with neurobiological evidence pointing to an involvement of reward-related brain areas in trust behavior (Delgado et al., 2005; Fehr & Camerer, 2007; King-Casas et al., 2005) as well as with recent research proposing an expressive in addition to an instrumental component of trust, assuming that individuals also consider the immediate "gains" (i.e., positive feelings) associated with the trusting act itself in their decision to trust (Dunning et al., 2014, 2012; Dunning & Fetchenhauer, 2010, 2011).

Specifically, individuals high in Extraversion should perceive social interactions as particularly rewarding per se (e.g., Ashton, Lee, & Paunonen, 2002; Denissen & Penke, 2008; Fishman, Ng, & Bellugi, 2011; Lucas, Diener, Grob, Suh, & Shao, 2000; Lucas & Diener, 2001; Pavot, Diener, & Fujita, 1990) and therefore be highly motivated to approach such interactions (Depue & Collins, 1999; Gray, 1970). In turn, extraverts might be more willing to trust others than introverts as a result of anticipating a large gain from trusting. Correspondingly, findings point to a positive relation between Extraversion and the willingness to trust (Evans & Revelle, 2008; Hiraishi, Yamagata, Shikishima, & Ando, 2008; Swope, Cadigan, Schmitt, & Shupp, 2008; Thielmann & Hilbig, 2014). Note, however, that the above reasoning presupposes a truly social situation involving interpersonal interaction. Particularly in the context of trust among strangers this may rarely be the case because many situations involve a limited amount of social interaction (e.g., online purchase). In turn, the effect of reward sensitivity is restricted to situations involving a high social component. Besides, given that most people should be motivated to avoid betrayal (as implied by a general tendency toward betrayal aversion; e.g., Bohnet & Zeckhauser, 2004), betrayal sensitivity should be the stronger predictor of trust behavior as compared to reward sensitivity. For these reasons, we primarily focus on betrayal sensitivity as a determinant of outcome evaluation in situations of trust among strangers.

The Big Picture: The Interplay of Trust Determinants

So far, we have outlined how the trusting choice of making oneself dependent on another can be understood as a function of three determinants: (I) attitudes toward risky prospects (i.e., risk aversion and loss aversion), (II) trustworthiness expectations (based on trust cues, prior trust experiences, and/or social projection), and (III) betrayal sensitivity. Assuming that the decision to trust is conceptually comparable to a decision under risk, risk aversion and loss aversion should essentially mediate the effects of trustworthiness expectations and betrayal sensitivity, respectively. Specifically, trustworthiness expectations influence the probability

[12] Unlike the (multiple) findings indicating betrayal aversion, there is also some evidence pointing to the opposite, namely a higher willingness to take a risk in the trust game compared to a risk game (Fetchenhauer & Dunning, 2009, 2012). However, in these studies, a participant's decision to trust in the trust game, but *not* her decision to take a risk in the risk game, determined the outcome of another person—which might in turn explain the deviating results.

[13] Note that, although similar in name, betrayal sensitivity has to be differentiated from rejection sensitivity, which denotes an individual's disposition to anxiously expect, readily perceive, and intensely respond to rejection (Downey & Feldman, 1996). By contrast, betrayal sensitivity does not involve specific expectations toward betrayal or a particular readiness to perceive betrayal, but rather specifically captures individuals' evaluation of losses due to betrayal (compared to losses attributable to nature).

input which is processed based on an individual's risk aversion. In turn, betrayal sensitivity influences the outcome input which is processed based on an individual's loss aversion (cf. Figure 1). In other words, (I) risk aversion and loss aversion essentially operate on (II) the subjective probability of betrayal (the result of the formation of trustworthiness expectations) and (III) the subjective outcome utility (the result of outcome evaluation as potentially distorted through betrayal sensitivity), respectively.

Although this combined view suggests that all three determinants make a unique contribution to the decision to trust, trustworthiness expectations seem especially important given that trust behavior is *based upon* some expectation of whether the trustee will act in a benevolent fashion. That is, it is arguably a necessary condition of trust that the trustor expects (i.e., subjectively estimates) at least a larger-than-zero probability of trust appreciation (cf. Giffin, 1967). In case a trustor expects a 0% probability that the trustee acts trustworthily (e.g., due to the presence of suspicious trust cues), the trustor should never trust, regardless of how small the potential loss (and how large the potential gain) of trusting might be. In effect, risk aversion, loss aversion, and betrayal sensitivity will thus be irrelevant.

In turn, whenever the estimated probability of trust appreciation exceeds zero, individuals' levels of risk aversion, loss aversion, and betrayal sensitivity might come into play. More specifically, these components may be particularly influential whenever a trustor has neither a very positive, nor a very negative expectation about the interaction partner's trustworthiness; that is, in situations in which the subjective chance of trust appreciation equals the subjective chance of trust betrayal (i.e., both approximate 50%). Indeed, evidence suggests that risk aversion is especially relevant for trust behavior when individuals have a rather vague trustworthiness expectation. For example, risk aversion has been found to be most predictive of trust behavior in individuals with neither a strong prosocial nor a strong proself SVO (Kanagaretnam et al., 2009) and thus in individuals who will not arrive at a strong expectation through social projection. Moreover, increasing the subjective confidence in trustworthiness expectations by providing specific information about an interaction partner (i.e., sex, preferences, appearance, ethnicity) eliminated the otherwise observed relation between risk aversion and trust behavior (Eckel & Wilson, 2004). Similarly, information about an interaction partner's likely back transfer reduced the (albeit small) impact of risk aversion on trust behavior (Houser et al., 2010). In sum, these findings suggest that the impact of risk aversion on trust behavior depends on the degree of uncertainty and risk expressed in the trustworthiness expectations. Arguably, the same should apply to loss aversion and betrayal sensitivity, though no empirical evidence for or against this conjecture is currently available.

Furthermore, it seems likely that *actual* decisions to trust are not universally and equally contingent on all trust determinants. If, for example, the trustor anticipates a highly positive outcome from trusting, this anticipated gain might outweigh the potential losses, especially in individuals low in loss aversion. As a consequence, the individual might place great emphasis on the resulting positive utility and essentially neglect the probabilities of trust appreciation versus trust betrayal in her decision to trust. In turn, dispositional risk aversion (which processes these probabilities) might only play a minor role for the particular decision to trust. The same might hold if the utility associated with trusting is highly negative, thus leading to distrust in the first place (again almost irrespective of dispositional risk aversion). Such interactions between determinants might hence also account for the inconsistent evidence relating dispositional risk aversion to trust behavior (in addition to measurement-related issues, as summarized above). In any case, the speculative nature of these arguments shows that further research will be necessary to clarify the potential interplay of different trust determinants (cf. Table 1). Even so, it seems at least reasonable that actual trust decisions will sometimes deviate from how trust decisions should be made based on consequentialist logic.

Models of Personality

As reasoned above, one of the key aims of the current review is to link the main components of trust to individual differences and personality traits. In particular, we focus on the links between models of basic personality structure and the three central components of trust summarized above so as to provide a theory-driven, comprehensive, and parsimonious account. Recall that we previously outlined three broad (combinations of) personality traits relevant for trust behavior: (I) anxiety and fear, explaining individual differences in risk and loss aversion, (II) fairness and honesty (i.e., trustworthiness), explaining individual differences in trustworthiness expectations through social projection, and (III) forgiveness, explaining individual differences in betrayal sensitivity. In the following, we will discuss whether and how these personality traits are covered in models of basic personality structure. To this end, we refer to two models that have been primarily considered in recent personality research on social behavior: the Five-Factor Model (Costa & McCrae, 1992; McCrae & Costa, 1985) and the HEXACO model of personality structure (Ashton & Lee, 2007; Lee & Ashton, 2004). Note that more specific traits that have been considered relevant predictors of social behavior—most prominently SVO (Van Lange, 1999)—can be subsumed under the broader traits conceptualized in the basic trait models.

The Five-Factor Model

As the name suggests, the Five-Factor Model of personality (FFM) entails five factors as the basic dimensions of individual differences: Neuroticism, Extraversion, Conscientiousness, Agreeableness, and Openness to Experience (Costa & McCrae, 1992; McCrae & Costa, 1985). With regard to the presented three-component framework—and the traits assumed to drive trust behavior (viz., anxiety/fear, trustworthiness, and forgiveness)—Neuroticism and Agreeableness are the two factors that should be primarily responsible for individual differences in trust behavior. In the following, we will briefly review evidence supporting this assumption.

As outlined above, trait anxiety and fear, respectively, should explain individual differences in risk aversion and loss aversion (cf. Figure 1) and thereby account for how expectations (i.e., subjective probabilities) and potential outcomes are processed and integrated. In the FFM, trait anxiety and fear are both included in the *anxiety* facet of Neuroticism (e.g., as assessed by the NEO Personality Inventory Revised, NEO-PI-R; Costa & McCrae, 1992) which, in turn, is referred to as "the sensitivity of a domain-general system to respond to environmental threats" (Denissen &

Penke, 2008, p. 1289). Correspondingly, several studies reported high correlations between Neuroticism and trait anxiety and/or fear (e.g., Doty, Japee, Ingvar, & Ungerleider, 2013; Gomez & Francis, 2003; Scheier, Carver, & Bridges, 1994). Moreover, high levels of Neuroticism have been identified as a risk factor for the development of anxiety disorders (Clark, Watson, & Mineka, 1994). Neuroticism should hence explain individual differences in trust behavior through its influence on trait anxiety and fear.

With regard to the second component of trust reviewed above, trustworthiness expectations, our summary of the literature suggests that social projection of an individual's own trustworthiness (i.e., fairness and honesty) will influence her expectations. In terms of the FFM, one's trustworthiness should in turn be covered by Agreeableness. From among the FFM dimensions, the latter has most consistently been related to fairness, with high Agreeableness predicting sharing in the dictator game (Baumert, Schlösser, & Schmitt, 2013; Becker, Deckers, Dohmen, Falk, & Kosse, 2012; Ben-Ner, Kong, & Putterman, 2004), nonexploitation in resource dilemmas (Koole, Jager, van den Berg, Vlek, & Hofstee, 2001) and public goods games (Volk, Thöni, & Ruigrok, 2011), cooperation in the prisoner's dilemma game (Pothos, Perry, Corr, Matthew, & Busemeyer, 2011), and trustworthiness in the trust game (Becker et al., 2012; Ben-Ner & Halldorsson, 2010)—although these effects are not entirely robust and null-findings have also been reported (Brocklebank, Lewis, & Bates, 2011; Evans & Revelle, 2008; Hirsh & Peterson, 2009; Kurzban & Houser, 2001; Lönnqvist, Verkasalo, & Walkowitz, 2011; Müller & Schwieren, 2012; for a recent review on the relation between behavior in economic games and basic personality traits see Zhao & Smillie, 2015). According to the presented framework, individuals high in Agreeableness might hence have rather optimistic expectations about another's trustworthiness (due to projecting their own tendency to be cooperative onto others), thus being more willing to trust than individuals low in Agreeableness. Indeed, Agreeableness specifically includes a *trust* facet, describing "the tendency to attribute benevolent intent to others" (Costa, McCrae, & Dye, 1991, p. 888)—thus essentially capturing the essence of trustworthiness expectations. Similarly, trait honesty—the second dimension of trustworthiness—is covered by the *straightforwardness* facet of Agreeableness as measured by the NEO-PI-R (Ashton et al., 2000; Costa et al., 1991), thus further supporting the notion that Agreeableness should account for individual differences in trustworthiness.

Third and finally, evidence suggests that forgiveness—the trait assumed to drive betrayal sensitivity—should be best explained by low Neuroticism and high Agreeableness. Specifically, low Neuroticism and high Agreeableness have consistently been associated with trait forgiveness and forgiveness-related behavior (Ashton, Paunonen, Helmes, & Jackson, 1998; J. W. Berry, Worthington, O'Connor, Parrott, & Wade, 2005; Brose, Rye, Lutz-Zois, & Ross, 2005; Jensen-Campbell & Graziano, 2001; Walker & Gorsuch, 2002; T.-W. Wang, 2008). In turn, high Neuroticism and low Agreeableness have been associated with vengefulness (Maltby et al., 2008; McCullough, Bellah, Kilpatrick, & Johnson, 2001; McCullough & Hoyt, 2002)—an expression of low forgiveness (e.g., McCullough et al., 2001). Thus, Neuroticism and Agreeableness should explain individual differences in trust behavior, in this case through their influence on trait forgiveness which arguably drives individuals' betrayal sensitivity.

Concluding from the evidence sketched so far, the FFM would account for individual differences in trust behavior primarily through two dimensions, namely Neuroticism and Agreeableness. With regard to Agreeableness, prior evidence directly supports this idea, showing that high Agreeableness has a positive effect on investments in the trust game (Becker et al., 2012; Ben-Ner & Halldorsson, 2010; Evans & Revelle, 2008; Mikolajczak et al., 2010; Zhao & Smillie, 2015) which could—in line with the presented framework—be particularly traced back to the *trust* and *straightforwardness* facets of Agreeableness (Müller & Schwieren, 2012). With regard to Neuroticism, however, evidence is rather scarce with only one study supporting the proposed negative relation between Neuroticism (specifically the *anxiety* facet) and trust (Müller & Schwieren, 2012) and another study hinting at a trend for said relation (Ben-Ner & Halldorsson, 2010). However, only a few studies have predicted trust behavior from the perspective of the FFM, and thus further research is needed to critically test the proposed associations between trust behavior and Neuroticism as well as Agreeableness.

The HEXACO Model

As an extension and slight variation of the FFM, the HEXACO model of personality structure (Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to Experience) has recently been proposed based on lexical studies across various languages, supporting a six-factor structure of basic personality traits (Ashton & Lee, 2007; Lee & Ashton, 2004). From a theoretical point of view, the HEXACO model specifically proposes three dimensions as underlying factors of prosocial behavior (Ashton, Lee, & De Vries, 2014; Ashton & Lee, 2007) and thus of trust: Emotionality, Honesty-Humility, and Agreeableness. We will outline how each of these three HEXACO dimensions particularly covers one of the proposed personality aspects underlying trust behavior (i.e., anxiety/fear, trustworthiness, and forgiveness) in what follows.

To begin with, anxiety and fear are both included in the Emotionality factor of the HEXACO model, a variant of FFM-Neuroticism. In particular, Emotionality contains one specific facet for each of these traits, with the *anxiety* facet capturing the tendency to worry and the *fearfulness* facet capturing the tendency to experience fear (Lee & Ashton, 2004). By definition, high levels of Emotionality are linked to "decreased opportunities for gains that are [. . .] associated with risks" (Ashton & Lee, 2007, p. 156). Emotionality can hence be assumed to account for individual differences in trust behavior based on its effect on trait anxiety and fear.

With regard to trustworthiness, the second trait component of trust proposed in the present framework, the HEXACO model clearly points to the newly introduced Honesty-Humility factor as the primary predictor of fairness and honesty. More specifically, Honesty-Humility represents "the tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation" (Ashton & Lee, 2007, p. 156). Correspondingly, high levels of Honesty-Humility have consistently been related to fairness and nonexploitation in economic games (Baumert et al., 2013; Hilbig, Glöckner, & Zettler, 2014; Hilbig et al., 2012; Hilbig, Thielmann, Hepp, Klein, & Zettler, 2015; Hilbig, Thielmann, Wührl, &

Zettler, 2015; Hilbig, Zettler, Leist, & Heydasch, 2013; Thielmann, Hilbig, & Niedtfeld, 2014; Thielmann & Hilbig, 2014; Zettler, Hilbig, & Heydasch, 2013) as well as to prosocial SVO (Ackermann, Fleiß, & Murphy, in press; Hilbig et al., 2014; Hilbig, Zettler, Moshagen, & Heydasch, 2013; Hilbig & Zettler, 2009). Low levels of Honesty-Humility, in contrast, have been associated with drive for money, material goods, and power (Ashton & Lee, 2008b; Lee et al., 2013) as well as with crime and delinquency (Ashton & Lee, 2008b; Cohen, Panter, Turan, Morse, & Kim, 2014; De Vries & van Gelder, 2013; Dunlop, Morrison, Koenig, & Silcox, 2012; Lee, Ashton, & De Vries, 2005; Van Gelder & De Vries, 2012; for an overview see Ashton & Lee, 2008a). Besides fairness, Honesty-Humility has—as its name suggests—also been linked to honesty-related behavior. Specifically, research found that low levels of Honesty-Humility predicted cheating (Hershfield, Cohen, & Thompson, 2012; Hilbig & Zettler, 2015), immoral behavior (Hilbig, Moshagen, & Zettler, 2015), and unethical business practices (Ashton & Lee, 2008b; Cohen et al., 2014; Lee, Ashton, Morrison, Cordery, & Dunlop, 2008; Marcus, Lee, & Ashton, 2007). Accounting for both fairness and honesty, the Honesty-Humility factor should hence cover the trustworthiness aspect of trust behavior. Supporting this reasoning, Honesty-Humility was significantly related to trustworthiness expectations, mediated by fairness as assessed in the dictator game (Thielmann & Hilbig, 2014).

Finally, trait forgiveness—the proposed source of individual differences in betrayal sensitivity—is covered by HEXACO-Agreeableness.[14] More specifically, HEXACO-Agreeableness represents "the tendency to be forgiving and tolerant of others, in the sense of cooperating with others even when one might be suffering exploitation by them" (Ashton & Lee, 2007, p. 156). Correspondingly, high levels of HEXACO-Agreeableness have been associated with trait forgiveness (Shepherd & Belicki, 2008) as well as with the willingness to refrain from retaliatory actions when others act in an uncooperative or unfair manner (Hilbig, Zettler, Leist, et al., 2013; Thielmann et al., 2014). In turn, low levels of HEXACO-Agreeableness have been linked to revenge intentions (Lee & Ashton, 2012; K. E. Sheppard & Boon, 2012) and readiness to retaliate (Perugini, Gallucci, Presaghi, & Ercolani, 2003). Overall, HEXACO-Agreeableness should explain individual differences in trust behavior through its influence on trait forgiveness.

To conclude, the HEXACO model would account for individual differences in trust behavior based on three basic traits, namely Emotionality, Honesty-Humility, and Agreeableness. Indeed, each of the three proposed characteristics underlying trust behavior (i.e., anxiety/fear, trustworthiness, and forgiveness) is—at least in theory—uniquely covered by one single personality trait of the HEXACO model. That is, unlike the FFM, the HEXACO model provides a more clear-cut distinction between the underlying traits of trust behavior. This will allow for stricter empirical tests at a higher resolution. For example, rather than predicting that some combination of Agreeableness and Neuroticism will influence trust behavior through trait forgiveness and betrayal sensitivity, respectively, the HEXACO model makes the bold assumption that HEXACO-Agreeableness alone would be responsible for this mechanism. However, actual evidence on the relation between trust behavior and the HEXACO dimensions is still missing (cf. Zhao & Smillie, 2015). We suggest that this is also a fruitful direction for future research.

However, as briefly sketched above, Extraversion might be another determinant of trust behavior (mediated through reward sensitivity) in situations involving a pronounced social component. In line with this notion, Extraversion has—in both the FFM and the HEXACO framework—been considered a pillar of the "interpersonal circle." That is, both models assume that Extraversion covers the intensity of social interaction, whereas Agreeableness (in the FFM) and Emotionality, Agreeableness, and Honesty-Humility (in the HEXACO) cover the quality of social interaction (Ashton et al., 2014; Ashton & Lee, 2007). Hence, the FFM as well as the HEXACO might account for the (social) reward-related aspect of trust through individual differences in Extraversion. A positive relation between Extraversion and trust is also implied by the positive secondary loading of the trust facet of FFM-Agreeableness (as measured via the NEO-PI-R; Costa et al., 1991) on Extraversion as well as by negative secondary loadings of the anxiety and fearfulness facets of HEXACO-Emotionality (Lee & Ashton, 2004) on Extraversion.

Summary and Conclusion

In the present work, we provide a broad and integrative review of interpersonal trust behavior and present a person-situation framework which organizes the extant findings on trust among strangers into a coherent structure. We conclude that trust among strangers can be explained through an interaction between features of the trust situation and a trustor's personality characteristics. In general, this view rests on the notion that trust behavior represents a risky choice to depend on another in a situation of uncertainty (cf. Boon & Holmes, 1991). That is, conceptually, the decision to trust is similar to a decision under risk with more or less obvious outcomes (i.e., potential gain and loss), and mostly unknown probabilities of these outcomes—because of the inherent uncertainty about an interaction partner's trustworthiness.

Based on this general approach, we distill three core components of interpersonal trust behavior from the extant literature: (I) attitudes toward risky prospects, (II) trustworthiness expectations, and (III) betrayal sensitivity (cf. Figure 1). First, in line with the decision making literature, we propose that trust behavior involves two attitudes toward risky prospects, one referring to risk-taking as a function of probabilities (i.e., risk aversion) and the other referring to risk-taking as a function of outcomes (i.e., loss aversion)—both, in turn, influenced by an individual's trait anxiety and fear. Second, given that the probabilities of gain versus loss (i.e., trust appreciation vs. trust betrayal) are, by definition, unknown in a trust situation (e.g., Yamagishi & Yamagishi, 1994), trustors might use three sources of information to come up with a trustworthiness expectation which provides the probability-related input for the decision to trust: (a) trust cues, (b) prior trust experiences, and (c) social projection. Whereas trust cues and prior trust experiences

[14] HEXACO-Agreeableness represents a slightly rotated version of FFM-Agreeableness, including, for example, anger-related traits that are, in the FFM, summarized in the Neuroticism factor. In the HEXACO model, trait forgiveness is hence uniquely ascribed to Agreeableness whereas in the FFM, it is covered by a combination of Agreeableness and Neuroticism (Ashton et al., 2014).

mainly refer to the specific trust situation, social projection refers to the personality aspect underlying trustworthiness expectations. That is, it reflects an individual's own trustworthiness (i.e., fairness and honesty) which is, in turn, projected onto another (e.g., Krueger et al., 2012). Third and finally, the risk associated with trusting results from another person's betrayal rather than from bad luck (cf. Bohnet & Zeckhauser, 2004). Thus, betrayal sensitivity is introduced as a determinant of trust-related loss evaluation that is driven by trait forgiveness and assumed to affect the outcome-related input of the decision to trust.

In general, the present work approaches interpersonal trust behavior from both a situational and an individual difference perspective. With regard to the latter (i.e., the personality aspect of trust behavior), we extend earlier notions of interpersonal trust by specifying three distinct personality traits (i.e., anxiety/fear, trustworthiness, and forgiveness) as primary sources of interindividual variation. These three traits, in turn, are well captured in models of basic personality structure, specifically in the Five-Factor Model (Costa & McCrae, 1992; McCrae & Costa, 1985) and the HEXACO model (Ashton & Lee, 2007; Lee & Ashton, 2004). Thus, precise predictions can be derived and tested beyond existing accounts that merely considered a single trait of trust propensity (e.g., Mayer et al., 1995). At the same time, the framework details different aspects of the trust situation (e.g., another's trustworthiness, temptation to betray, availability of potential sanctions) which are likely to influence trust behavior independent of (or in some interaction with) the proposed personality aspects.

In addition, the present work addresses several contradicting findings and unsolved issues reported in the trust literature. First, the review points to a reasonable explanation for people's seemingly irrational tendency to trust despite rather pessimistic expectations about others' trustworthiness (Dufwenberg & Gneezy, 2000; Evans & Krueger, 2014; Fetchenhauer & Dunning, 2009): Trust behavior is not solely driven by expectations; rather—as soon as an individual expects a minimum chance of trust appreciation—other factors affect the decision to trust as well (i.e., risk aversion, loss aversion, and betrayal sensitivity). Hence, individuals might have good reasons to trust despite expecting a relatively low probability of trust appreciation. Also, the review explicitly traces the frequently observed (yet insufficiently explained) relation between trust and trustworthiness (e.g., Colquitt, Scott, & LePine, 2007; see also early work by Wrightsman, 1964) to social projection of a trustor's own trustworthiness onto interaction partners. Finally, it provides reasonable explanations for the inconsistent evidence regarding the relation between trust behavior and risk aversion. On the one hand, the appropriate assessment and conceptualization of dispositional risk aversion is still unclear. Thus, inconsistent findings seem to be likely attributable to the reliance on different and partly suboptimal measures of risk aversion which, at the same time, apply to different domains of risk aversion. On the other hand, the influence of dispositional risk aversion on trust behavior might be diminished under certain circumstances as a result of the interplay between the different trust determinants. As sketched above, it seems likely that—depending on the weights individuals assign to the different aspects of the situation (i.e., gains and losses with corresponding probabilities of occurrence)—the aspects may outweigh each other, thus causing one determinant to be the primary driver of the specific trust decision. This implies that dispositional risk aversion, for example, will drive the decision to trust only weakly if individuals place great emphasis on the potential outcomes of trusting. Altogether, our review and corresponding framework hence indicates that—even in light of the inconsistent extant evidence—risk aversion should be maintained as a reasonable predictor of trust behavior.

In going beyond a summary of and common structure for extant literature, the present work also provides clear-cut hypotheses, implications, and specific open questions for future research (cf. Table 1). In particular, it is our hope that this review might trigger more systematic attempts to study the relation between trust behavior and (basic) personality traits—an issue that has, in our view, gained too little attention in previous work. Similarly, the framework might inspire research on the interaction between (and integration of) the different determinants underlying trust decisions rather than focusing on each component in isolation. This might help clarify inconsistent evidence reported in the extant literature. Overall, the review and distilled framework might encourage bringing together different areas of research (e.g., behavioral economics, judgment and decision making, social psychology, and personality psychology), thereby fostering a symbiosis between these (typically—and unfortunately—rather secluded) behavioral sciences as has been called for (cf. Rhodewalt, 2008).

Moreover, the specification of three trait components underlying trust might help clarify sex- and age-related differences in trust behavior. Regarding the former, the typically observed higher levels of trait anxiety among women (Costa, Terracciano, & McCrae, 2001; Feingold, 1994)—and the correspondingly elevated levels of risk aversion (Eckel & Grossman, 2008), loss aversion (e.g., Brooks & Zank, 2005; Rau, 2014; Schmidt & Traub, 2002), and Emotionality (Moshagen, Hilbig, & Zettler, 2014)—imply that women would show a reduced tendency to trust as compared to men. By contrast, however, women's tendency to be more forgiving than men (A. J. Miller, Worthington, & McDaniel, 2008) suggests the opposite, namely an increased tendency to trust. Finally, equal levels of cooperativeness across the sexes (Balliet, Li, Macfarlan, & Van Vugt, 2011) hint at similar trustworthiness expectations (based on social projection) and thus a lack of sex differences in trust. Altogether, sex differences in trait anxiety, forgiveness, and cooperativeness hence lead to conflicting predictions on sex differences in trust behavior. However, these inconsistencies actually correspond to the mixed extant evidence on sex differences in trust—as reported in a recent literature review (Croson & Gneezy, 2009)—and indeed make sense within the current framework. Specifically, because of the proposed interplay of trust determinants, the most decisive determinant in a given situation might drive the presence (vs. absence) and direction (higher levels in men vs. women) of sex differences in trust behavior. Besides, regarding age-related differences in trust behavior, lower levels of risk aversion in the social domain in older compared to younger adults (Bonem, Ellsworth, & Gonzalez, 2015) suggest a higher willingness to trust in elderly people. A similar prediction can be derived from more optimistic trustworthiness expectations (i.e., higher trust propensity) in older adults (Li & Fung, 2013; Poulin & Haase, in press). However, contrary to these predictions, the (albeit scarce) extant evidence on trust behavior across ages hints at constant trust levels in different adult age groups (Rieger & Mata, 2015; Sutter & Kocher, 2007). In any

case, more conclusive data on age- (and sex-) related differences in trust behavior (and their underlying trait sources) is needed.

It should also be noted that the approach taken herein is well-suited for future expansions to other types of trust, for example trust among familiar agents (e.g., colleagues, friends, family and the like). In interactions with familiar others, one could, for instance, assume that trustworthiness expectations are primarily based on prior trust experiences with the specific interaction partner in similar situations—in line with interdependence theory (Holmes, 2002). In consequence, these expectations should be clearer, thus reducing uncertainty and fostering predictability (cf. Holmes & Rempel, 1989) and, in turn, diminishing the impact of risk aversion on trust behavior. In contrast, the effect of betrayal sensitivity could arguably be enhanced because a loss resulting from betrayal by a (well-)known person may well be even more severe than a loss resulting from betrayal by an unknown person (cf. Rotenberg, 2010). Thus, comparing trust in strangers to trust in familiar agents, one could make the strong prediction that the relative impact of risk aversion will be greater in the former than in the latter—and vice versa for betrayal sensitivity (and, by implication, loss aversion). Of course, additional factors might play a role for the decision to trust a familiar interaction partner, for example reciprocity considerations (Ferrin, Bligh, & Kohles, 2007) or a motivation to communicate positive relational signals (Six, Nooteboom, & Hoogendoorn, 2010; see also Simpson, 2007). In any case, although trust among familiar agents is beyond the scope of this review we are confident that it can be integrated into the framework proposed—or a variant of the latter.

In conclusion, the present review provides an integrative summary on trust behavior among unfamiliar agents, spanning several hundred articles. Scholars from various fields of research agree on the vital significance of trust for us as social beings, but also point to the apparent irrationality of trusting unknown others. Correspondingly, the study of trust in general—and of trust among strangers in particular—has gained broad attention in research across all social and economic sciences and beyond. So far, however, there has been little integration of the diverse literature on trust within a coherent structure. To this end, we present a person–situation integration framework of interpersonal trust behavior, embedding the (specific) literature on trust across fields in the (more general) literature on behavioral economics, social psychology, and personality psychology. Thereby, the present work contributes to an overall understanding of trust among strangers by uncovering the underlying determinants of trust behavior from both a situational and a personality perspective.

Specifically, the review highlights that people might decide to trust for various reasons: On the one hand, diverse features of the trust situation (i.e., trustee characteristics, temptation to betray, availability of potential sanctions) might signal the likelihood of encountering a trustworthy (vs. untrustworthy) interaction partner; on the other hand, different personality characteristics (i.e., risk and loss aversion, trustworthiness, and betrayal sensitivity) might influence an individual's willingness to trust irrespective of (or in interaction with) the situation at hand. As such, the extracted framework specifically accounts for individual differences in trust behavior—an issue that has been clearly underemphasized in prior conceptualizations of trust. Most notably, however, the work presented herein provides a broad summary of the extensive literature on trust and may thus offer a theoretical foundation for future research.

References

Ackermann, K. A., Fleiß, J., & Murphy, R. O. (in press). Reciprocity as an individual difference. *The Journal of Conflict Resolution*. Advance online publication. http://dx.doi.org/10.1177/0022002714541854

Aimone, J. A., & Houser, D. (2011). Beneficial betrayal aversion. *PLoS ONE, 6,* e17725. http://dx.doi.org/10.1371/journal.pone.0017725

Aimone, J. A., & Houser, D. (2012). What you don't know won't hurt you: A laboratory analysis of betrayal aversion. *Experimental Economics, 15,* 571–588. http://dx.doi.org/10.1007/s10683-012-9314-z

Aimone, J. A., & Houser, D. (2013). Harnessing the benefits of betrayal aversion. *Journal of Economic Behavior & Organization, 89,* 1–8. http://dx.doi.org/10.1016/j.jebo.2013.02.001

Albert, M., Güth, W., Kirchler, E., & Maciejovsky, B. (2007). Are we nice(r) to nice(r) people?–An experimental analysis. *Experimental Economics, 10,* 53–69. http://dx.doi.org/10.1007/s10683-006-9131-3

Altmann, S., Dohmen, T., & Wibral, M. (2008). Do the reciprocal trust less? *Economics Letters, 99,* 454–457. http://dx.doi.org/10.1016/j.econlet.2007.09.012

Anh, H. T., Pereira, L. M., & Santos, F. C. (2011). Intention recognition promotes the emergence of cooperation. *Adaptive Behavior, 19,* 264–279. http://dx.doi.org/10.1177/1059712311410896

Arneson, R. J. (1982). The principle of fairness and free-rider problems. *Ethics, 92,* 616. http://dx.doi.org/10.1086/292379

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70,* 1–70. http://dx.doi.org/10.1037/h0093718

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics, 9,* 193–208. http://dx.doi.org/10.1007/s10683-006-9122-4

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11,* 150–166. http://dx.doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008a). The HEXACO model of personality structure and the importance of the H factor. *Social and Personality Psychology Compass, 2,* 1952–1962. http://dx.doi.org/10.1111/j.1751-9004.2008.00134.x

Ashton, M. C., & Lee, K. (2008b). The prediction of Honesty-Humility–related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality, 42,* 1216–1228. http://dx.doi.org/10.1016/j.jrp.2008.03.006

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18,* 139–152. http://dx.doi.org/10.1177/1088868314523838

Ashton, M. C., Lee, K., & Paunonen, S. V. (2002). What is the central feature of extraversion? Social attention versus reward sensitivity. *Journal of Personality and Social Psychology, 83,* 245–252. http://dx.doi.org/10.1037/0022-3514.83.1.245

Ashton, M. C., Lee, K., & Son, C. (2000). Honesty as the sixth factor of personality: Correlations with Machiavellianism, primary psychopathy, and social adroitness. *European Journal of Personality, 14,* 359–368. http://dx.doi.org/10.1002/1099-0984(200007/08)14:4<359::AID-PER382>3.0.CO;2-Y

Ashton, M. C., Paunonen, S. V., Helmes, E., & Jackson, D. N. (1998). Kin altruism, reciprocal altruism, and the Big Five personality factors. *Evolution and Human Behavior, 19,* 243–255. http://dx.doi.org/10.1016/S1090-5138(98)00009-9

Bacharach, M., & Gambetta, D. (2001). Trust in signs. In K. S. Cook (Ed.), *Trust in society* (pp. 148–184). New York, NY: Russell Sage Foundation.

Baier, A. (1986). Trust and antitrust. *Ethics, 96,* 231–260. http://dx.doi.org/10.1086/292745

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin, 137,* 881–909. http://dx.doi.org/10.1037/a0025354

Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin, 137,* 594–615. http://dx.doi.org/10.1037/a0023489

Balliet, D., Parks, C. D., & Joireman, J. A. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations, 12,* 533–547. http://dx.doi.org/10.1177/1368430209105040

Balliet, D., & Van Lange, P. A. M. (2012). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin, 139,* 1090–1112. http://dx.doi.org/10.1037/a0030939

Balliet, D., & Van Lange, P. A. M. (2013). Trust, punishment, and cooperation across 18 societies: A meta-analysis. *Perspectives on Psychological Science, 8,* 363–379. http://dx.doi.org/10.1177/1745691613488533

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin, 140,* 1556–1581. http://dx.doi.org/10.1037/a0037737

Barclay, P. (2004). Trustworthiness and competitive altruism can also solve the "tragedy of the commons." *Evolution and Human Behavior, 25,* 209–220. http://dx.doi.org/10.1016/j.evolhumbehav.2004.04.002

Baron, J. (1998). Trust: Beliefs and morality. In A. Ben-Ner & L. Putterman (Eds.), *Economics, values, and organization* (pp. 408–418). Cambridge, England: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139174855.017

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323–370. http://dx.doi.org/10.1037/1089-2680.5.4.323

Baumert, A., Schlösser, T. M., & Schmitt, M. (2013). Economic games: A performance-based assessment of fairness and altruism. *European Journal of Psychological Assessment*. Advance online publication.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58,* 639–650. http://dx.doi.org/10.1016/j.neuron.2008.04.009

Becker, A., Deckers, T., Dohmen, T., Falk, A., & Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics, 4,* 453–478. http://dx.doi.org/10.1146/annurev-economics-080511-110922

Bellemare, C., & Kroger, S. (2007). On representative social capital. *European Economic Review, 51,* 183–202. http://dx.doi.org/10.1016/j.euroecorev.2006.03.006

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology, 31,* 64–79. http://dx.doi.org/10.1016/j.joep.2009.10.001

Ben-Ner, A., Kong, F., & Putterman, L. (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology, 25,* 581–589. http://dx.doi.org/10.1016/S0167-4870(03)00065-5

Ben-Ner, A., & Putterman, L. (2009). Trust, communication and contracts: An experiment. *Journal of Economic Behavior & Organization, 70(1–2): 10,* 6–121. http://dx.doi.org/10.1016/j.jebo.2009.01.011

Bente, G., Baptist, O., & Leuschner, H. (2012). To buy or not to buy: Influence of seller photos and reputation on buyer trust and purchase behavior. *International Journal of Human-Computer Studies, 70,* 1–13. http://dx.doi.org/10.1016/j.ijhcs.2011.08.005

Berg, J., Dickhaut, J., & McCabe, K. A. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10,* 122–142. http://dx.doi.org/10.1006/game.1995.1027

Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition, 8,* 343–361. http://dx.doi.org/10.1521/soco.1990.8.4.343

Berry, J. W., Worthington, E. L. J., Jr., O'Connor, L. E., Parrott, L., III, & Wade, N. G. (2005). Forgivingness, vengeful rumination, and affective traits. *Journal of Personality, 73,* 183–226. http://dx.doi.org/10.1111/j.1467-6494.2004.00308.x

Betsch, T., & Glöckner, A. (2010). Intuition in judgment and decision making: Extensive thinking without effort. *Psychological Inquiry, 21,* 279–294. http://dx.doi.org/10.1080/1047840X.2010.517737

Betts, L. R., Rotenberg, K. J., & Trueman, M. (2009). The Early Childhood Generalized Trust Belief Scale. *Early Childhood Research Quarterly, 24,* 175–185. http://dx.doi.org/10.1016/j.ecresq.2008.10.002

Bibby, P. A., & Ferguson, E. (2011). The ability to process emotional information predicts loss aversion. *Personality and Individual Differences, 51,* 263–266. http://dx.doi.org/10.1016/j.paid.2010.05.001

Bicchieri, C., Duffy, J., & Tolle, G. (2004). Trust among strangers. *Philosophy of Science, 71,* 286–319. http://dx.doi.org/10.1086/381411

Bigoni, M., Bortolotti, S., Casari, M., & Gambetta, D. (2013). It takes two to cheat: An experiment on derived trust. *European Economic Review, 64,* 129–146. http://dx.doi.org/10.1016/j.euroecorev.2013.08.009

Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., & Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. *Personality and Individual Differences, 69,* 56–61. http://dx.doi.org/10.1016/j.paid.2014.05.012

Blair, M. M., & Stout, L. A. (2001). Trust, trustworthiness, and the behavioral foundations of corporate law. *University of Pennsylvania Law Review, 149,* 1735–1810. http://dx.doi.org/10.2307/3312898

Blais, A.-R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1,* 33–47.

Blomqvist, K. (1997). The many faces of trust. *Scandinavian Journal of Management, 13,* 271–286. http://dx.doi.org/10.1016/S0956-5221(97)84644-1

Boero, R., Bravo, G., Castellani, M., & Squazzoni, F. (2009). Reputational cues in repeated trust games. *The Journal of Socio-Economics, 38,* 871–877. http://dx.doi.org/10.1016/j.socec.2009.05.004

Bohnet, I., & Baytelman, Y. (2007). Institutions and trust: Implications for preferences, beliefs and behavior. *Rationality and Society, 19,* 99–135. http://dx.doi.org/10.1177/1043463107075110

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *The American Economic Review, 98,* 294–310. http://dx.doi.org/10.1257/aer.98.1.294

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). Trust and the reference points for trustworthiness in Gulf and Western countries. *The Quarterly Journal of Economics, 125,* 811–828. http://dx.doi.org/10.1162/qjec.2010.125.2.811

Bohnet, I., & Huck, S. (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *The American Economic Review, 94,* 362–366. http://dx.doi.org/10.1257/0002828041301506

Bohnet, I., & Meier, S. (2005). *Deciding to distrust. Public policy discussion papers, Federal Reserve Bank of Boston, No, 05:4*. http://dx.doi.org/10.2139/ssrn.839225

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization, 55,* 467–484. http://dx.doi.org/10.1016/j.jebo.2003.11.004

Bolton, G. E., Katok, E., & Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science, 50,* 1587–1602. http://dx.doi.org/10.1287/mnsc.1030.0199

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of equity, reciprocity, and competition. *The American Economic Review, 90,* 166–193. http://dx.doi.org/10.1257/aer.90.1.166

Bond, C. F., Jr., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology, 15,* 523–534. http://dx.doi.org/10.1207/s15324834basp1504_8

Bonem, E. M., Ellsworth, P. C., & Gonzalez, R. (2015). Age differences in risk: Perceptions, intentions and domains. [Advance online publication]. *Journal of Behavioral Decision Making,* n/a. http://dx.doi.org/10.1002/bdm.1848

Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General, 142,* 143–150. http://dx.doi.org/10.1037/a0028930

Boon, S. D., & Holmes, J. G. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In R. A. Hinde & J. Groebel (Eds.), *Cooperation and prosocial behavior* (pp. 190–211). Cambridge, England: Cambridge University Press.

Boone, R. T., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior, 27,* 163–182. http://dx.doi.org/10.1023/A:1025341931128

Bracht, J., & Feltovich, N. (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics, 93,* 1036–1044. http://dx.doi.org/10.1016/j.jpubeco.2009.06.004

Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes* (pp. 215–232). New York, NY: Psychology Press.

Brocklebank, S., Lewis, G. J., & Bates, T. C. (2011). Personality accounts for stable preferences and expectations across a range of simple games. *Personality and Individual Differences, 51,* 881–886. http://dx.doi.org/10.1016/j.paid.2011.07.007

Brooks, P., & Zank, H. (2005). Loss averse behavior. *Journal of Risk and Uncertainty, 31,* 301–325. http://dx.doi.org/10.1007/s11166-005-5105-7

Brose, L. A., Rye, M. S., Lutz-Zois, C., & Ross, S. R. (2005). Forgiveness and personality traits. *Personality and Individual Differences, 39,* 35–46. http://dx.doi.org/10.1016/j.paid.2004.11.001

Brunswik, E. (1952). *The conceptual framework of psychology*. Oxford, UK: University Chicago Press.

Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization, 51,* 195–216. http://dx.doi.org/10.1016/S0167-2681(02)00093-8

Butler, G., & Mathews, A. (1987). Anticipatory anxiety and risk perception. *Cognitive Therapy and Research, 11,* 551–565. http://dx.doi.org/10.1007/BF01183858

Cabon-Dhersin, M.-L., & Ramani, S. V. (2007). Opportunism, trust and cooperation: A game theoretic approach with heterogeneous agents. *Rationality and Society, 19,* 203–228. http://dx.doi.org/10.1177/1043463107077391

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. New York, NY: Russell Sage Foundation.

Camerer, C. F. (2005). Three cheers–psychological, theoretical, empirical–for loss aversion. *Journal of Marketing Research, 42,* 129–133. http://dx.doi.org/10.1509/jmkr.42.2.129.62286

Camerer, C. F., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica, 56,* 1–36. http://dx.doi.org/10.2307/1911840

Campellone, T. R., & Kring, A. M. (2013). Who do you trust? The impact of facial emotion and behaviour on decision making. *Cognition and Emotion, 27,* 603–620. http://dx.doi.org/10.1080/02699931.2012.726608

Chang, L. J., Doll, B. B., van 't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61,* 87–105. http://dx.doi.org/10.1016/j.cogpsych.2010.03.001

Charness, G., Cobo-Reyes, R., & Jimenez, N. (2008). An investment game with third-party intervention. *Journal of Economic Behavior & Organization, 68,* 18–28. http://dx.doi.org/10.1016/j.jebo.2008.02.006

Charness, G., Du, N., & Yang, C.-L. (2011). Trust and trustworthiness reputations in an investment game. *Games and Economic Behavior, 72,* 361–375. http://dx.doi.org/10.1016/j.geb.2010.09.002

Chaudhuri, A., & Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal, 73,* 959–985.

Chen, X.-P., Pillutla, M. M., & Yao, X. (2009). Unintended consequences of cooperation inducing and maintaining mechanisms in public goods dilemmas: Sanctions and moral appeals. *Group Processes & Intergroup Relations, 12,* 241–255. http://dx.doi.org/10.1177/1368430208098783

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103,* 103–116. http://dx.doi.org/10.1037/0021-843X.103.1.103

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology, 107,* 943–963. http://dx.doi.org/10.1037/a0037245

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA: Belknap Press.

Colquitt, J. A., Scott, B. A., & LePine, J. A. (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,* 909–927. http://dx.doi.org/10.1037/0021-9010.92.4.909

Corcos, A., Pannequin, F., & Bourgeois-Gironde, S. (2012). Aversions to trust. *Recherches Economiques de Louvain/Louvain. Economic Review, 78,* 115–134.

Coricelli, G., Gonzalez Morales, L., & Mahlstedt, A. (2006). The investment game with asymmetric information. *Metroeconomica, 57,* 13–30. http://dx.doi.org/10.1111/j.1467-999X.2006.00231.x

Corriveau, K., & Harris, P. L. (2010). Young children's trust in what other people say. In K. J. Rotenberg (Ed.), *Interpersonal trust during childhood and adolescence* (pp. 87–109). New York, NY: New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511750946.005

Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12,* 887–898. http://dx.doi.org/10.1016/0191-8869(91)90177-D

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and NEO Five–Factor Inventory (NEO–FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/0022-3514.81.2.322

Courtois, P., & Tazdaït, T. (2012). Learning to trust strangers: An evolutionary perspective. *Journal of Evolutionary Economics, 22,* 367–383. http://dx.doi.org/10.1007/s00191-011-0247-z

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior, 46,* 260–281. http://dx.doi.org/10.1016/S0899-8256(03)00119-2

Crosetto, P., & Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty, 47,* 31–65. http://dx.doi.org/10.1007/s11166-013-9170-z

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature, 47,* 448–474. http://dx.doi.org/10.1257/jel.47.2.448

Currall, S. C., & Judge, T. A. (1995). Measuring trust between organizational boundary role persons. *Organizational Behavior and Human Decision Processes, 64,* 151–170. http://dx.doi.org/10.1006/obhd.1995.1097

Das, T. K., & Teng, B.-S. (2004). The risk-based view of trust: A conceptual framework. *Journal of Business and Psychology, 19,* 85–116. http://dx.doi.org/10.1023/B:JOBU.0000040274.23551.1b

Dasgupta, P. (1988). Trust as a commodity. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (Vol. 4, pp. 49–72). Oxford, UK: Basil Blackwell.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology, 35,* 1–11. http://dx.doi.org/10.1037/0022-3514.35.1.1

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8,* 1611–1618. http://dx.doi.org/10.1038/nn1575

De Neys, W., Hopfensitz, A., & Bonnefon, J.-F. (2015). Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults. *Journal of Economic Psychology, 47,* 17–22. http://dx.doi.org/10.1016/j.joep.2015.01.002

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor Model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality, 42,* 1285–1302. http://dx.doi.org/10.1016/j.jrp.2008.04.002

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences, 22,* 491–517. http://dx.doi.org/10.1017/S0140525X99002046

Derks, J., Lee, N. C., & Krabbendam, L. (2014). Adolescent trust and trustworthiness: Role of gender and social value orientation. *Journal of Adolescence, 37,* 1379–1386. http://dx.doi.org/10.1016/j.adolescence.2014.09.014

Desmet, P. T. M., De Cremer, D., & van Dijk, E. (2011). Trust recovery following voluntary or forced financial compensations in the trust game: The role of trait forgiveness. *Personality and Individual Differences, 51,* 267–273. http://dx.doi.org/10.1016/j.paid.2010.05.027

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science, 23,* 1549–1556. http://dx.doi.org/10.1177/0956797612448793

Deutsch, M. (1958). Trust and suspicion. *The Journal of Conflict Resolution, 2,* 265–279. http://dx.doi.org/10.1177/002200275800200401

Deutsch, M. (1960). Trust, trustworthiness, and the F scale. *The Journal of Abnormal and Social Psychology, 61,* 138–140. http://dx.doi.org/10.1037/h0046501

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1962* (pp. 275–320). Lincoln, NE: University of Nebraska Press.

De Vries, R. E., & van Gelder, J.-L. (2013). Tales of two self-control scales: Relations with Five-Factor and HEXACO traits. *Personality and Individual Differences, 54,* 756–760. http://dx.doi.org/10.1016/j.paid.2012.12.023

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2011). The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*. Advance online publication.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association, 9,* 522–550. http://dx.doi.org/10.1111/j.1542-4774.2011.01015.x

Doty, T. J., Japee, S., Ingvar, M., & Ungerleider, L. G. (2013). Fearful face detection sensitivity in healthy adults correlates with anxiety-related traits. *Emotion, 13,* 183–188. http://dx.doi.org/10.1037/a0031373

Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology, 70,* 1327–1343. http://dx.doi.org/10.1037/0022-3514.70.6.1327

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior, 30,* 163–182. http://dx.doi.org/10.1006/game.1999.0715

Dunlop, P. D., Morrison, D. L., Koenig, J., & Silcox, B. (2012). Comparing the Eysenck and HEXACO models of personality in the prediction of adult delinquency. *European Journal of Personality, 26,* 194–202. http://dx.doi.org/10.1002/per.824

Dunn, J. (1988). Trust and political agency. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 73–93). Oxford, UK: Basil Blackwell.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology, 88,* 736–748. http://dx.doi.org/10.1037/0022-3514.88.5.736

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107,* 122–141. http://dx.doi.org/10.1037/a0036673

Dunning, D. A., & Fetchenhauer, D. (2010). Trust as an expressive rather than an instrumental act. In S. Thye & E. Lawler (Eds.), *Advances in group processes* (Vol. 27, pp. 97–127). New York, NY: Emerald. http://dx.doi.org/10.1108/S0882-6145(2010)0000027007

Dunning, D. A., & Fetchenhauer, D. (2011). Understanding the psychology of trust. In D. A. Dunning (Ed.), *Social motivation* (pp. 147–169). New York, NY: Psychology Press.

Dunning, D. A., Fetchenhauer, D., & Schlösser, T. M. (2012). Trust as a social and emotional act: Noneconomic considerations in trust behavior. *Journal of Economic Psychology, 33,* 686–694. http://dx.doi.org/10.1016/j.joep.2011.09.005

Eckel, C. C., & Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results, 1,* 1061–1073. http://dx.doi.org/10.1016/S1574-0722(07)00057-1

Eckel, C. C., & Wilson, R. K. (2003). The human face of game theory: Trust and reciprocity in sequential games. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 245–274). New York, NY: Russell Sage Foundation.

Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization, 55,* 447–465. http://dx.doi.org/10.1016/j.jebo.2003.11.003

Eek, D., & Biel, A. (2003). The interplay between greed, efficiency, and fairness in public-goods dilemmas. *Social Justice Research, 16,* 195–215. http://dx.doi.org/10.1023/A:1025984611796

Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports, 3,* 1047. http://dx.doi.org/10.1038/srep01047

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics, 75,* 643–669. http://dx.doi.org/10.2307/1884324

Etang, A., Fielding, D., & Knowles, S. (2011). Does trust extend beyond the village? Experimental trust and social distance in Cameroon. *Experimental Economics, 14,* 15–35. http://dx.doi.org/10.1007/s10683-010-9255-3

Evans, A. M., Athenstaedt, U., & Krueger, J. I. (2013). The development of trust and altruism during childhood. *Journal of Economic Psychology, 36,* 82–95. http://dx.doi.org/10.1016/j.joep.2013.02.010

Evans, A. M., & Krueger, J. I. (2009). The psychology (and economics) of trust. *Social and Personality Psychology Compass, 3,* 1003–1017. http://dx.doi.org/10.1111/j.1751-9004.2009.00232.x

Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk and perspective-taking. *Journal of Experimental Social Psychology, 47,* 171–177. http://dx.doi.org/10.1016/j.jesp.2010.08.007

Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust. *Judgment and Decision Making, 9,* 90–103.

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality, 42,* 1585–1593. http://dx.doi.org/10.1016/j.jrp.2008.07.011

Fang, C., Kimbrough, S. O., Pace, S., Valluri, A., & Zheng, Z. (2002). On adaptive emergence of trust behavior in the game of stag hunt. *Group Decision and Negotiation, 11,* 449–467. http://dx.doi.org/10.1023/A:1020639132471

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association, 7,* 235–266. http://dx.doi.org/10.1162/JEEA.2009.7.2-3.235

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: The neural circuitry of social preferences. *Trends in Cognitive Sciences, 11,* 419–427. http://dx.doi.org/10.1016/j.tics.2007.09.002

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114,* 817–868. http://dx.doi.org/10.1162/003355399556151

Fehrler, S., & Przepiorka, W. (2013). Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts. *Evolution and Human Behavior, 34,* 139–145. http://dx.doi.org/10.1016/j.evolhumbehav.2012.11.005

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116,* 429–456. http://dx.doi.org/10.1037/0033-2909.116.3.429

Ferrin, D. L., Bligh, M. C., & Kohles, J. C. (2007). Can I trust you to trust me? A theory of trust, monitoring, and cooperation in interpersonal and intergroup relationships. *Group & Organization Management, 32,* 465–499. http://dx.doi.org/10.1177/1059601106293960

Fetchenhauer, D., & Dunning, D. A. (2009). Do people trust too much or too little? *Journal of Economic Psychology, 30,* 263–276. http://dx.doi.org/10.1016/j.joep.2008.04.006

Fetchenhauer, D., & Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science, 21,* 189–193. http://dx.doi.org/10.1177/0956797609358586

Fetchenhauer, D., & Dunning, D. A. (2012). Betrayal aversion versus principled trustfulness—How to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior & Organization, 81,* 534–541. http://dx.doi.org/10.1016/j.jebo.2011.07.017

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology, 82,* 956–974. http://dx.doi.org/10.1037/0022-3514.82.6.956

Fishman, I., Ng, R., & Bellugi, U. (2011). Do extraverts process social stimuli differently from introverts? *Cognitive Neuroscience, 2,* 67–73. http://dx.doi.org/10.1080/17588928.2010.527434

Fleeson, W., & Leicht, C. (2006). On delineating and integrating the study of variability and stability in personality psychology: Interpersonal trust as illustration. *Journal of Research in Personality, 40,* 5–20. http://dx.doi.org/10.1016/j.jrp.2005.08.004

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science, 20,* 419–422. http://dx.doi.org/10.1111/j.1467-9280.2009.02312.x

Funder, D. C. (2008). Persons, situations, and person-situation interactions. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 568–580). New York, NY: Guilford Press.

Gächter, S., Johnson, E. J., & Herrmann, A. (2010). *Individual-level loss aversion in riskless and risky choices*. CeDEx Discussion Paper Series, 2010–20, Center for Decision Research & Experimental Economics.

Gambetta, D. (1988). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–237). Oxford, UK: Basil Blackwell.

Gambetta, D., & Przepiorka, W. (2014). Natural and strategic generosity as signals of trustworthiness. *PLoS ONE, 9,* e97533.

Giffin, K. (1967). The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin, 68,* 104–120. http://dx.doi.org/10.1037/h0024833

Gigerenzer, G. (1998). Surrogates for theories. *Theory & Psychology, 8,* 195–204. http://dx.doi.org/10.1177/0959354398082006

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669. http://dx.doi.org/10.1037/0033-295X.103.4.650

Gill, H., Boies, K., Finegan, J. E., & McNally, J. (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,* 287–302. http://dx.doi.org/10.1007/s10869-004-2229-8

Gilliam, D. A., & Voss, K. (2013). A proposed procedure for construct definition in marketing. *European Journal of Marketing, 47,* 5–26. http://dx.doi.org/10.1108/03090561311285439

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24,* 153–172. http://dx.doi.org/10.1016/S1090-5138(02)00157-5

Giorgetta, C., Grecucci, A., Zuanon, S., Perini, L., Balestrieri, M., Bonini, N., . . . Brambilla, P. (2012). Reduced risk-taking behavior as a trait feature of anxiety. *Emotion, 12,* 1373–1383. http://dx.doi.org/10.1037/a0029119

Glanville, J. L., & Paxton, P. (2007). How do we learn to trust? A confirmatory tetrad analysis of the sources of generalized trust. *Social Psychology Quarterly, 70,* 230–242. http://dx.doi.org/10.1177/019027250707000303

Glöckner, A., & Hilbig, B. E. (2012). Risk is relative: Risk aversion yields cooperation rather than defection in cooperation-friendly environments. *Psychonomic Bulletin & Review, 19,* 546–553. http://dx.doi.org/10.3758/s13423-012-0224-z

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking & Reasoning, 16,* 1–25. http://dx.doi.org/10.1080/13546780903395748

Golembiewski, R. T., & McConkie, M. (1975). The centrality of interpersonal trust in group processes. In C. L. Cooper (Ed.), *Theories of group processes* (pp. 131–185). London, UK: Wiley.

Gollwitzer, M., Rothmund, T., Alt, B., & Jekel, M. (2012). Victim sensitivity and the accuracy of social judgments. *Personality and Social Psychology Bulletin, 38,* 975–984. http://dx.doi.org/10.1177/0146167212440887

Gollwitzer, M., Rothmund, T., Pfeiffer, A., & Ensenbach, C. (2009). Why and when justice sensitivity leads to pro- and antisocial behavior. *Journal of Research in Personality, 43,* 999–1005. http://dx.doi.org/10.1016/j.jrp.2009.07.003

Gollwitzer, M., Rothmund, T., & Süssenbach, P. (2013). The Sensitivity to Mean Intentions (SeMI) model: Basic assumptions, recent findings, and potential avenues for future research. *Social and Personality Psychology Compass, 7,* 415–426. http://dx.doi.org/10.1111/spc3.12041

Gomez, R., & Francis, L. M. (2003). Generalised anxiety disorder: Relationships with Eysenck's, Gray's and Newman's theories. *Personality and Individual Differences, 34,* 3–17. http://dx.doi.org/10.1016/S0191-8869(02)00020-X

Goto, S. G. (1996). To trust or not to trust: Situational and dispositional determinants. *Social Behavior and Personality, 24,* 119–131. http://dx.doi.org/10.2224/sbp.1996.24.2.119

Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy, 8,* 249–266. http://dx.doi.org/10.1016/0005-7967(70)90069-0

Güney, S., & Newell, B. R. (2013). Fairness overrides reputation: The importance of fairness considerations in altruistic cooperation. *Frontiers in Human Neuroscience, 7,* 252. http://dx.doi.org/10.3389/fnhum.2013.00252

Hardin, R. (2002). *Trust and trustworthiness*. New York, NY: Russell Sage Foundation.

Harris, A. J. L., Corner, A., & Hahn, U. (2009). Estimating the probability of negative events. *Cognition, 110,* 51–64. http://dx.doi.org/10.1016/j.cognition.2008.10.006

Harris, P. L. (2007). Trust. *Developmental Science, 10,* 135–138. http://dx.doi.org/10.1111/j.1467-7687.2007.00575.x

Hershfield, H. E., Cohen, T. R., & Thompson, L. (2012). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. *Organizational Behavior and Human Decision Processes, 117,* 298–310. http://dx.doi.org/10.1016/j.obhdp.2011.11.002

Hilbig, B. E., Glöckner, A., & Zettler, I. (2014). Personality and prosocial behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology, 107,* 529–539. http://dx.doi.org/10.1037/a0036074

Hilbig, B. E., Moshagen, M., & Zettler, I. (2015). Truth will out: Linking personality, morality, and honesty through indirect questioning. *Social Psychological and Personality Science, 6,* 140–147. http://dx.doi.org/10.1177/1948550614553640

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S., & Zettler, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality, 55,* 46–50. http://dx.doi.org/10.1016/j.jrp.2014.12.004

Hilbig, B. E., Thielmann, I., Wührl, J., & Zettler, I. (2015). From Honesty–Humility to fair behavior—Benevolence or a (blind) fairness norm? *Personality and Individual Differences, 80,* 91–95. http://dx.doi.org/10.1016/j.paid.2015.02.017

Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty-Humility, social value orientations, and economic behavior. *Journal of Research in Personality, 43,* 516–519. http://dx.doi.org/10.1016/j.jrp.2009.01.003

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality, 57,* 72–88. http://dx.doi.org/10.1016/j.jrp.2015.04.003

Hilbig, B. E., Zettler, I., & Heydasch, T. (2012). Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional Honesty-Humility. *European Journal of Personality, 26,* 245–254. http://dx.doi.org/10.1002/per.830

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences, 54,* 598–603. http://dx.doi.org/10.1016/j.paid.2012.11.008

Hilbig, B. E., Zettler, I., Moshagen, M., & Heydasch, T. (2013). Tracing the path from personality—via cooperativeness—to conservation. *European Journal of Personality, 27,* 319–327. http://dx.doi.org/10.1002/per.1856

Hill, C. A., & O'Hara, E. A. (2006). A cognitive theory of trust. *Washington University Law Review, 84,* 1717–1796.

Hiraishi, K., Yamagata, S., Shikishima, C., & Ando, J. (2008). Maintenance of genetic variation in personality through control of mental mechanisms: A test of trust, extraversion, and agreeableness. *Evolution and Human Behavior, 29,* 79–85. http://dx.doi.org/10.1016/j.evolhumbehav.2007.07.004

Hirsh, J. B., & Peterson, J. B. (2009). Extraversion, neuroticism, and the prisoner's dilemma. *Personality and Individual Differences, 46,* 254–256. http://dx.doi.org/10.1016/j.paid.2008.10.006

Hodges, B. H. (2014). Rethinking conformity and imitation: Divergence, convergence, and social understanding. *Frontiers in Psychology, 5,* 726. http://dx.doi.org/10.3389/fpsyg.2014.00726

Hodges, B. H., Meagher, B. R., Norton, D. J., McBain, R., & Sroubek, A. (2014). Speaking from ignorance: Not agreeing with others we believe are correct. *Journal of Personality and Social Psychology, 106,* 218–234. http://dx.doi.org/10.1037/a0034662

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review, 114,* 733–758. http://dx.doi.org/10.1037/0033-295X.114.3.733

Holm, H. J., & Danielson, A. (2005). Tropic trust versus Nordic trust: Experimental evidence from Tanzania and Sweden. *The Economic Journal, 115,* 505–532. http://dx.doi.org/10.1111/j.1468-0297.2005.00998.x

Holm, H. J., & Nystedt, P. (2008). Trust in surveys and games–A methodological contribution on the influence of money and location. *Journal of Economic Psychology, 29,* 522–542. http://dx.doi.org/10.1016/j.joep.2007.07.010

Holmes, J. G. (2002). Interpersonal expectations as the building blocks of social cognition: An interdependence theory perspective. *Personal Relationships, 9,* 1–26. http://dx.doi.org/10.1111/1475-6811.00001

Holmes, J. G., & Rempel, J. K. (1989). Trust in close relationships. In C. Hendrick (Ed.), *Close relationships* (pp. 187–220). Thousand Oaks, CA: Sage.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review, 92,* 1644–1655. http://dx.doi.org/10.1257/000282802762024700

Hong, K., & Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology, 28,* 197–213. http://dx.doi.org/10.1016/j.joep.2006.06.003

Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *The Academy of Management Review, 20,* 379–403.

Houser, D., Schunk, D., & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization, 74,* 72–81. http://dx.doi.org/10.1016/j.jebo.2010.01.002

James, H. S. J. (2002). On the reliability of trusting. *Rationality and Society, 14,* 229–256. http://dx.doi.org/10.1177/1043463102014002004

Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality, 69,* 323–362. http://dx.doi.org/10.1111/1467-6494.00148

Johansson-Stenman, O., Mahmud, M., & Martinsson, P. (2013). Trust, trust games and stated trust: Evidence from rural Bangladesh. *Journal of Economic Behavior & Organization, 95,* 286–298. http://dx.doi.org/10.1016/j.jebo.2011.06.022

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32,* 865–889. http://dx.doi.org/10.1016/j.joep.2011.05.007

Johnson-George, C., & Swap, W. C. (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,* 1306–1317. http://dx.doi.org/10.1037/0022-3514.43.6.1306

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *The Journal of Business, 59,* S285–S300. http://dx.doi.org/10.1086/296367

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511809477

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291. http://dx.doi.org/10.2307/1914185

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341–350. http://dx.doi.org/10.1037/0003-066X.39.4.341

Kamas, L., & Preston, A. (2012). Distributive and reciprocal fairness: What can we learn from the heterogeneity of social preferences? *Journal of Economic Psychology, 33,* 538–553. http://dx.doi.org/10.1016/j.joep.2011.12.003

Kanagaretnam, K., Mestelman, S., Nainar, K., & Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. *Journal of Economic Psychology, 30,* 368–380. http://dx.doi.org/10.1016/j.joep.2008.12.003

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review, 95,* 1688–1699. http://dx.doi.org/10.1257/000282805775014407

Kee, H. W., & Knox, R. E. (1970). Conceptual and methodological considerations in the study of trust and suspicion. *Journal of Conflict Resolution, 14,* 357–366. http://dx.doi.org/10.1177/002200277001400307

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An atlas of interpersonal situations*. New York, NY: Cambridge University Press.

Kelley, H. H., & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology, 16,* 66–91. http://dx.doi.org/10.1037/h0029849

Keser, C. (2003). Experimental games for the design of reputation management systems. *IBM Systems Journal, 42,* 498–506. http://dx.doi.org/10.1147/sj.423.0498

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308,* 78–83. http://dx.doi.org/10.1126/science.1108062

Kirsch, P., Esslinger, C., Chen, Q., Mier, D., Lis, S., Siddhanti, S., . . . Meyer-Lindenberg, A. (2005). Oxytocin modulates neural circuitry for social cognition and fear in humans. *The Journal of Neuroscience, 25,* 11489–11493. http://dx.doi.org/10.1523/JNEUROSCI.3984-05.2005

Kiyonari, T., & Yamagishi, T. (1999). A comparative study of trust and trustworthiness using the game of enthronement. *The Japanese Journal of Social Psychology, 15,* 100–109.

Kiyonari, T., Yamagishi, T., Cook, K. S., & Cheshire, C. (2006). Does trust beget trustworthiness? Trust and trustworthiness in two games and two cultures: A research note. *Social Psychology Quarterly, 69,* 270–283. http://dx.doi.org/10.1177/019027250606900304

Klapwijk, A., & Van Lange, P. A. M. (2009). Promoting cooperation and trust in "noisy" situations: The power of generosity. *Journal of Personality and Social Psychology, 96,* 83–103. http://dx.doi.org/10.1037/a0012823

Koenig, M. A., & Harris, P. L. (2005). The role of social cognition in early trust. *Trends in Cognitive Sciences, 9,* 457–459. http://dx.doi.org/10.1016/j.tics.2005.08.006

Kollock, P. (1999). The production of trust in online markets. In S. R. Thye, E. J. Lawler, M. W. Macy, & H. A. Walker (Eds.), *Advances in group processes* (Vol. 16, pp. 99–123). Greenwich, CT: JAI Press.

Koole, S. L., Jager, W., van den Berg, A. E., Vlek, C. A. J., & Hofstee, W. K. B. (2001). On the social nature of personality: Effects of Extraversion, Agreeableness, and feedback about collective resource use on cooperation in a resource dilemma. *Personality and Social Psychology Bulletin, 27,* 289–301. http://dx.doi.org/10.1177/0146167201273003

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435,* 673–676. http://dx.doi.org/10.1038/nature03701

Kramer, R. M. (2010). Dilemmas and doubts: How decision-makers cope with interdependence and uncertainty. In R. M. Kramer, A. E. Tenbrunsel, & M. H. Bazerman (Eds.), *Social decision making: Social dilemmas, social values, and ethical judgments* (pp. 117–143). New York, NY: Routledge/Taylor & Francis Group.

Kreps, D. M. (1990). Corporate culture and economic theory. In J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on positive political economy* (pp. 90–143). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1017/CBO9780511571657.006

Krueger, J. I. (2007). From social projection to social behaviour. *European Review of Social Psychology, 18,* 1–35. http://dx.doi.org/10.1080/10463280701284645

Krueger, J. I., & Acevedo, M. (2005). Social projection and the psychology of choice. In M. D. Alicke, D. A. Dunning, & J. I.. Krueger (Eds.), *The self in social judgment* (pp. 17–41). New York, NY: Psychology Press.

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning. *The American Journal of Psychology, 120,* 593–618.

Krueger, J. I., DiDonato, T. E., & Freestone, D. (2012). Social projection can solve social dilemmas. *Psychological Inquiry, 23,* 1–27. http://dx.doi.org/10.1080/1047840X.2012.641167

Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence to social norms. *Negotiation and Conflict Management Research, 1,* 31–52. http://dx.doi.org/10.1111/j.1750-4716.2007.00003.x

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7,* 730–735. http://dx.doi.org/10.1037/1528-3542.7.4.730

Kugler, T., Connolly, T., & Kausel, E. E. (2009). The effect of consequential thinking on trust game behavior. *Journal of Behavioral Decision Making, 22,* 101–119. http://dx.doi.org/10.1002/bdm.614

Kurzban, R., & Houser, D. (2001). Individual differences in cooperation in a circular public goods game. *European Journal of Personality, 15*(1,SpecIssue), S37–S52. http://dx.doi.org/10.1002/per.420

Landrum, A. R., Eaves, B. S. J., Jr., & Shafto, P. (2015). Learning to trust and trusting to learn: A theoretical framework. *Trends in Cognitive Sciences, 19,* 109–111. http://dx.doi.org/10.1016/j.tics.2014.12.007

Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences, 31,* 215–226. http://dx.doi.org/10.1016/S0191-8869(00)00130-6

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39,* 329–358. http://dx.doi.org/10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2012). Getting mad and getting even: Agreeableness and Honesty-Humility as predictors of revenge intentions. *Personality and Individual Differences, 52,* 596–600. http://dx.doi.org/10.1016/j.paid.2011.12.004

Lee, K., Ashton, M. C., & De Vries, R. E. (2005). Predicting workplace delinquency and integrity with the HEXACO and Five-Factor models of personality structure. *Human Performance, 18,* 179–197. http://dx.doi.org/10.1207/s15327043hup1802_4

Lee, K., Ashton, M. C., Morrison, D. L., Cordery, J., & Dunlop, P. D. (2008). Predicting integrity with the HEXACO personality model: Use of self- and observer reports. *Journal of Occupational and Organizational Psychology, 81,* 147–167. http://dx.doi.org/10.1348/096317907X195175

Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the Dark Triad and Honesty–Humility. *European Journal of Personality, 27,* 169–184. http://dx.doi.org/10.1002/per.1860

Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality traits beyond the big five: Are they within the HEXACO space? *Journal of Personality, 73,* 1437–1463. http://dx.doi.org/10.1111/j.1467-6494.2005.00354.x

Lehmann-Waffenschmidt, M., & Leipold, K. (2011). Determinants of trust. *Nervenheilkunde: Zeitschrift für Interdisziplinaere Fortbildung, 30,* 401–409.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8,* 75–84. http://dx.doi.org/10.1037/1076-898X.8.2.75

Lenton, P., & Mosley, P. (2011). Incentivising trust. *Journal of Economic Psychology, 32,* 890–897. http://dx.doi.org/10.1016/j.joep.2011.07.005

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81,* 146–159. http://dx.doi.org/10.1037/0022-3514.81.1.146

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *The Academy of Management Review, 23,* 438–458.

Li, T., & Fung, H. H. (2013). Age differences in trust: An investigation across 38 countries. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 68,* 347–355.

Little, A. C., Jones, B. C., DeBruine, L. M., & Dunbar, R. I. M. (2013). Accuracy in discrimination of self-reported cooperators using static facial information. *Personality and Individual Differences, 54,* 507–512. http://dx.doi.org/10.1016/j.paid.2012.10.018

Lönnqvist, J.-E., Verkasalo, M., & Walkowitz, G. (2011). It pays to pay—Big Five personality influences on co-operative behaviour in an incentivized and hypothetical prisoner's dilemma game. *Personality and Individual Differences, 50,* 300–304. http://dx.doi.org/10.1016/j.paid.2010.10.009

Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G., & Wichardt, P. C. (2010). *Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison*. CGS Working Paper, Vol. 2, No. 3, Cologne Graduate School in Management, Economics and Social Sciences. http://dx.doi.org/10.2139/ssrn.1556647

López-Pérez, R. (2012). The power of words: A model of honesty and fairness. *Journal of Economic Psychology, 33,* 642–658. http://dx.doi.org/10.1016/j.joep.2011.12.004

Lorian, C. N., & Grisham, J. R. (2010). The safety bias: Risk-avoidance and social anxiety pathology. *Behaviour Change, 27,* 29–41. http://dx.doi.org/10.1375/bech.27.1.29

Lorian, C. N., & Grisham, J. R. (2011). Clinical implications of risk aversion: An online study of risk-avoidance and treatment utilization in pathological anxiety. *Journal of Anxiety Disorders, 25,* 840–848. http://dx.doi.org/10.1016/j.janxdis.2011.04.008

Lorian, C. N., Mahoney, A., & Grisham, J. R. (2012). Playing it safe: An examination of risk-avoidance in an anxious treatment-seeking sample. *Journal of Affective Disorders, 141,* 63–71. http://dx.doi.org/10.1016/j.jad.2012.02.021

Lucas, R. E., & Diener, E. (2001). Understanding extraverts' enjoyment of social situations: The importance of pleasantness. *Journal of Personality and Social Psychology, 81,* 343–356. http://dx.doi.org/10.1037/0022-3514.81.2.343

Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology, 79,* 452–468. http://dx.doi.org/10.1037/0022-3514.79.3.452

Luhmann, N. (1988). Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 94–107). Oxford, UK: Basil Blackwell.

Macko, A., Malawski, M., & Tyszka, T. (2014). Belief in others' trustworthiness and trusting behaviour. *Polish Psychological Bulletin, 45,* 43–51. http://dx.doi.org/10.2478/ppb-2014-0007

Macy, M. W., & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review, 63,* 638–660. http://dx.doi.org/10.2307/2657332

Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes, 94,* 61–73. http://dx.doi.org/10.1016/j.obhdp.2004.03.001

Maltby, J., Wood, A. M., Day, L., Kon, T. W. H., Colley, A., & Linley, P. A. (2008). Personality predictors of levels of forgiveness two and a half years after the transgression. *Journal of Research in Personality, 42,* 1088–1094. http://dx.doi.org/10.1016/j.jrp.2007.12.008

Manapat, M. L., Nowak, M. A., & Rand, D. G. (2013). Information, irrationality, and the evolution of trust. *Journal of Economic Behavior & Organization, 90*(Suppl), S57–S75. http://dx.doi.org/10.1016/j.jebo.2012.10.018

Manapat, M. L., & Rand, D. G. (2012). Delayed and inconsistent information and the evolution of trust. *Dynamic Games and Applications, 2,* 401–410. http://dx.doi.org/10.1007/s13235-012-0055-6

Maner, J. K., & Gerend, M. A. (2007). Motivationally selective risk judgments: Do fear and curiosity boost the boons or the banes? *Organizational Behavior and Human Decision Processes, 103,* 256–267. http://dx.doi.org/10.1016/j.obhdp.2006.08.002

Maner, J. K., Richey, J. A., Cromer, K., Mallott, M., Lejuez, C. W., Joiner, T. E., & Schmidt, N. B. (2007). Dispositional anxiety and risk-avoidant decision-making. *Personality and Individual Differences, 42,* 665–675. http://dx.doi.org/10.1016/j.paid.2006.08.016

Maner, J. K., & Schmidt, N. B. (2006). The role of risk avoidance in anxiety. *Behavior Therapy, 37,* 181–189. http://dx.doi.org/10.1016/j.beth.2005.11.003

Mansbridge, J. (1999). Altruistic trust. In M. Warren (Ed.), *Democracy and trust*. Cambridge, MA: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511659959.010

Manson, J. H., Gervais, M. M., & Kline, M. A. (2013). Defectors cannot be detected during "small talk" with strangers. *PLoS ONE, 8,* e82531. http://dx.doi.org/10.1371/journal.pone.0082531

Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition? *Personnel Psychology, 60,* 1–34. http://dx.doi.org/10.1111/j.1744-6570.2007.00063.x

Masuda, N., & Nakamura, M. (2012). Coevolution of trustful buyers and cooperative sellers in the trust game. *PLoS ONE, 7,* e44169. http://dx.doi.org/10.1371/journal.pone.0044169

Matthews, B. A., Kordonski, W. M., & Shimoff, E. (1983). Temptation and the maintenance of trust: Effects of bilateral punishment capability. *Journal of Conflict Resolution, 27,* 255–277. http://dx.doi.org/10.1177/0022002783027002003

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20,* 709–734.

McCarter, M. W., Rockmann, K. W., & Northcraft, G. B. (2010). Is it even worth it? The effect of loss prospects in the outcome distribution of a public goods dilemma. *Organizational Behavior and Human Decision Processes, 111,* 1–12. http://dx.doi.org/10.1016/j.obhdp.2009.06.003

McCrae, R. R., & Costa, P. T., Jr. (1985). Updating Norman's "Adequate Taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49,* 710–721. http://dx.doi.org/10.1037/0022-3514.49.3.710

McCullough, M. E., Bellah, C. G., Kilpatrick, S. D., & Johnson, J. L. (2001). Vengefulness: Relationships with forgiveness, rumination, well-being, and the Big Five. *Personality and Social Psychology Bulletin, 27,* 601–610. http://dx.doi.org/10.1177/0146167201275008

McCullough, M. E., & Hoyt, W. T. (2002). Transgression-related motivational dispositions: Personality substrates of forgiveness and their links to the Big Five. *Personality and Social Psychology Bulletin, 28,* 1556–1573. http://dx.doi.org/10.1177/014616702237583

McEvily, B., Radzevick, J. R., & Weber, R. A. (2012). Whom do you distrust and how much does it cost? An experiment on the measurement of trust. *Games and Economic Behavior, 74,* 285–298. http://dx.doi.org/10.1016/j.geb.2011.06.011

McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.), *Trust in cyber societies: Integrating the human and artificial perspectives* (pp. 27–54). Berlin, Germany: Springer. http://dx.doi.org/10.1007/3-540-45547-7_3

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *The Academy of Management Review, 23,* 473–490.

McNaughton, N. (2011). Trait anxiety, trait fear and emotionality: The perspective from non-human studies. *Personality and Individual Differences, 50,* 898–906. http://dx.doi.org/10.1016/j.paid.2010.07.011

Meertens, R. M., & Lion, R. (2008). Measuring an individual's tendency to take risks: The risk propensity scale. *Journal of Applied Social Psychology, 38,* 1506–1520. http://dx.doi.org/10.1111/j.1559-1816.2008.00357.x

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology, 4,* 1–25. http://dx.doi.org/10.1016/0022-1031(68)90046-2

Meyerson, D., Weick, K. E., & Kramer, R. M. (1996). Swift trust and temporary groups. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166–195). Thousand Oaks, CA: Sage http://dx.doi.org/10.4135/9781452243610.n9

Mikolajczak, M., Gross, J. J., Lane, A., Corneille, O., de Timary, P., & Luminet, O. (2010). Oxytocin makes people trusting, not gullible. *Psychological Science, 21,* 1072–1074. http://dx.doi.org/10.1177/0956797610377343

Miller, A. J., Worthington, E. L., Jr., & McDaniel, M. A. (2008). Gender and forgiveness: A meta–analytic review and research agenda. *Journal of Social and Clinical Psychology, 27,* 843–876. http://dx.doi.org/10.1521/jscp.2008.27.8.843

Miller, D. T. (1999). The norm of self-interest. *American Psychologist, 54,* 1053–1060. http://dx.doi.org/10.1037/0003-066X.54.12.1053

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology, 49,* 404–418. http://dx.doi.org/10.1037/a0029500

Moorman, C., Zaltman, G., & Deshpande, R. (1992). Relationships between providers and users of market research: The dynamics of trust within and between organizations. *Journal of Marketing Research, 29,* 314–328. http://dx.doi.org/10.2307/3172742

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorenstruktur, psychometrische Eigenschaften und Messinvarianz der deutschsprachigen Version des 60-Item HEXACO Persönlichkeitsinventars [Factor structure, psychometric properties, and measurement invariance of the German-language version of the 60-item HEXACO personality inventory]. *Diagnostica, 60,* 86–97. http://dx.doi.org/10.1026/0012-1924/a000112

Mulder, L. B., van Dijk, E., De Cremer, D., & Wilke, H. A. M. (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology, 42,* 147–162. http://dx.doi.org/10.1016/j.jesp.2005.03.002

Müller, J., & Schwieren, C. (2012). *What can the Big Five personality factors contribute to explain small-scale economic behavior?* Tinbergen Institute Discussion Paper No. 12–028/1. Amsterdam, The Netherlands: Tinbergen Institute. http://dx.doi.org/10.2139/ssrn.2029016

Murray, S. L., & Holmes, J. G. (2009). The architecture of interdependent minds: A Motivation-management theory of mutual responsiveness. *Psychological Review, 116,* 908–928. http://dx.doi.org/10.1037/a0017015

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin, 35,* 1661–1671. http://dx.doi.org/10.1177/0146167209346309

Newell, B. R., & Bröder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making, 3,* 195–204.

Oda, R., Naganawa, T., Yamauchi, S., Yamagata, N., & Matsumoto-Oda, A. (2009). Altruists are trusted based on non-verbal cues. *Biology Letters, 5,* 752–754. http://dx.doi.org/10.1098/rsbl.2009.0332

Orbell, J. M. (1993). Hamlet and the psychology of rational choice under uncertainty. *Rationality and Society, 5,* 127–140. http://dx.doi.org/10.1177/1043463193005001010

Orbell, J. M., & Dawes, R. M. (1991). A "cognitive miser" theory of cooperators' advantage. *The American Political Science Review, 85,* 515–528. http://dx.doi.org/10.2307/1963172

Pavot, W., Diener, E., & Fujita, F. (1990). Extraversion and happiness. *Personality and Individual Differences, 11,* 1299–1306. http://dx.doi.org/10.1016/0191-8869(90)90157-M

Payne, J. W. (1973). Alternative approaches to decision making under risk: Moments versus risk dimensions. *Psychological Bulletin, 80,* 439–453. http://dx.doi.org/10.1037/h0035260

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality, 17,* 251–283. http://dx.doi.org/10.1002/per.474

Platow, M. J., Foddy, M., Yamagishi, T., Lim, L., & Chow, A. (2012). Two experimental tests of trust in in-group strangers: The moderating role of common knowledge of group membership. *European Journal of Social Psychology, 42,* 30–35. http://dx.doi.org/10.1002/ejsp.852

Posten, A.-C., Ockenfels, A., & Mussweiler, T. (2014). How activating cognitive content shapes trust: A subliminal priming study. *Journal of Economic Psychology, 41,* 12–19. http://dx.doi.org/10.1016/j.joep.2013.04.002

Pothos, E. M., Perry, G., Corr, P. J., Matthew, M. R., & Busemeyer, J. R. (2011). Understanding cooperation in the prisoner's dilemma game. *Personality and Individual Differences, 51,* 210–215. http://dx.doi.org/10.1016/j.paid.2010.05.002

Poulin, M. J., & Haase, C. M. (in press). Growing to trust: Evidence that trust increases and becomes more important for well-being across the life span. *Social Psychological & Personality Science*. Advance online publication.

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology, 28,* 363–392. http://dx.doi.org/10.1146/annurev.ps.28.020177.002051

Przepiorka, W. (2013). Buyers pay for and sellers invest in a good reputation: More evidence from eBay. *The Journal of Socio-Economics, 42,* 31–42. http://dx.doi.org/10.1016/j.socec.2012.11.004

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review, 83,* 1281–1302.

Rau, H. A. (2014). The disposition effect and loss aversion: Do gender differences matter? *Economics Letters, 123,* 33–36. http://dx.doi.org/10.1016/j.econlet.2014.01.020

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49,* 95–112. http://dx.doi.org/10.1037/0022-3514.49.1.95

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *Advances in Applied Microeconomics, 11,* 127–157.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE, 7,* ArtID: E34293scheier. http://dx.doi.org/10.1371/journal.pone.0034293

Rhodewalt, F. (2008). Personality and social behavior: An overview. In F. Rhodewalt (Ed.), *Personality and social behavior* (pp. 1–8). New York, NY: Psychology Press.

Rieger, M., & Mata, R. (2015). On the generality of age differences in social and nonsocial decision making. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 70,* 200–214. http://dx.doi.org/10.1093/geronb/gbt088

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62,* 23–48. http://dx.doi.org/10.1146/annurev.psych.121208.131647

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly, 41,* 574–599. http://dx.doi.org/10.2307/2393868

Rosenberg, S., & Sedlak, A. (1972). Structural representations of implicit personality theory. *Advances in Experimental Social Psychology, 6,* 235–297.

Ross, W., & LaCroix, J. (1996). Multiple meanings of trust in negotiation theory and research: A literature review and integrative model. *International Journal of Conflict Management, 7,* 314–360. http://dx.doi.org/10.1108/eb022786

Rotenberg, K. J. (2010). The conceptualization of interpersonal trust: A basis, domain, and target framework. In K. J. Rotenberg (Ed.), *Interpersonal trust during childhood and adolescence* (pp. 8–27). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511750946.002

Rothmund, T., Gollwitzer, M., Bender, J., & Klimmt, C. (2015). Short- and long-term effects of video game violence on interpersonal trust. *Media Psychology, 18,* 106–133. http://dx.doi.org/10.1080/15213269.2013.841526

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35,* 651–665. http://dx.doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26,* 443–452. http://dx.doi.org/10.1037/h0031464

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35,* 1–7. http://dx.doi.org/10.1037/0003-066X.35.1.1

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. F. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23,* 393–404. http://dx.doi.org/10.5465/AMR.1998.926617

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5,* 296–320. http://dx.doi.org/10.1207/S15327957PSPR0504_2

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104,* 409–426. http://dx.doi.org/10.1037/a0031050

Sapienza, P., Toldra Simats, A., & Zingales, L. (2013). Understanding trust. *The Economic Journal, 123,* 1313–1332. http://dx.doi.org/10.1111/ecoj.12036

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22,* 617–640. http://dx.doi.org/10.1016/S0167-4870(01)00059-9

Schechter, L. (2007). Traditional trust measurement and the risk confound: An experiment in rural Paraguay. *Journal of Economic Behavior & Organization, 62,* 272–292. http://dx.doi.org/10.1016/j.jebo.2005.03.006

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67,* 1063–1078. http://dx.doi.org/10.1037/0022-3514.67.6.1063

Schlenker, B. R., Helm, B., & Tedeschi, J. T. (1973). The effects of personality and situational variables on behavioral trust. *Journal of Personality and Social Psychology, 25,* 419–427. http://dx.doi.org/10.1037/h0034088

Schmidt, U., & Traub, S. (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty, 25,* 233–249. http://dx.doi.org/10.1023/A:1020923921649

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science, 15,* 436–447. http://dx.doi.org/10.1111/j.1467-7687.2012.01135.x

Shelley, M. K. (1994). Individual differences in lottery evaluation models. *Organizational Behavior and Human Decision Processes, 60,* 206–230. http://dx.doi.org/10.1006/obhd.1994.1081

Shepherd, S., & Belicki, K. (2008). Trait forgiveness and traitedness within the HEXACO model of personality. *Personality and Individual Differences, 45,* 389–394. http://dx.doi.org/10.1016/j.paid.2008.05.011

Sheppard, B. H., & Sherman, D. M. (1998). The grammars of trust: A model and general implications. *The Academy of Management Review, 23,* 422–437.

Sheppard, K. E., & Boon, S. D. (2012). Predicting appraisals of romantic revenge: The roles of Honesty-Humility, Agreeableness, and vengefulness. *Personality and Individual Differences, 52,* 128–132. http://dx.doi.org/10.1016/j.paid.2011.09.014

Shinada, M., Yamagishi, T., Tanida, S., Takahashi, C., Inukai, K., Koizumi, M., . . . Hashimoto, H. (2010). [Accuracy of judgment about others' cooperative behavior: Effects of attractiveness and facial expressiveness]. *Japanese Journal of Phycology, 81,* 149–157. http://dx.doi.org/10.4992/jjpsy.81.149

Simpson, J. A. (2007). Foundations of interpersonal trust. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 587–607). New York, NY: Guilford Press.

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science, 4*(3), 367–392. http://dx.doi.org/10.1287/orsc.4.3.367

Six, F., Nooteboom, B., & Hoogendoorn, A. (2010). Actions that build interpersonal trust: A relational signalling perspective. *Review of Social Economy, 68,* 285–315. http://dx.doi.org/10.1080/00346760902756487

Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology, 78,* 1–18. http://dx.doi.org/10.1037/h0026468

Snijders, C., & Keren, G. (1999). Determinants of trust. In D. V. Budescu, I. Erev, & R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport* (pp. 355–385). Mahwah, NJ: Erlbaum Publishers.

Snijders, C., & Keren, G. (2001). Do you trust? Whom do you trust? When do you trust? In S. R. Thye (Ed.), *Advances in group processes* (Vol. 18, pp. 129–160). Bingley, UK: Emerald Group Publishing Limited. http://dx.doi.org/10.1016/S0882-6145(01)18006-9

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychological Science, 26,* 39–47. http://dx.doi.org/10.1177/0956797614554955

Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General, 141,* 715–727. http://dx.doi.org/10.1037/a0027950

Stiff, C. (2008). Are they bothered? How the opportunity to damage a partner's reputation influences giving behavior in a trust game. *The Journal of Social Psychology, 148,* 609–630. http://dx.doi.org/10.3200/SOCP.148.5.609-630

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21,* 349–354. http://dx.doi.org/10.1177/0956797610362647

Stouten, J., De Cremer, D., & van Dijk, E. (2006). Violating equality in social dilemmas: Emotional and retributive reactions as a function of trust, attribution, and honesty. *Personality and Social Psychology Bulletin, 32,* 894–906. http://dx.doi.org/10.1177/0146167206287538

Sutter, M., & Kocher, M. G. (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior, 59,* 364–382. http://dx.doi.org/10.1016/j.geb.2006.07.006

Swope, K. J., Cadigan, J., Schmitt, P. M., & Shupp, R. (2008). Personality preferences in laboratory economics experiments. *The Journal of Socio-Economics, 37,* 998–1009. http://dx.doi.org/10.1016/j.socec.2006.12.065

Sylwester, K., Lyons, M., Buchanan, C., Nettle, D., & Roberts, G. (2012). The role of Theory of Mind in assessing cooperative intentions. *Personality and Individual Differences, 52,* 113–117. http://dx.doi.org/10.1016/j.paid.2011.09.005

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality, 50,* 61–65. http://dx.doi.org/10.1016/j.jrp.2014.03.006

Thielmann, I., Hilbig, B. E., & Niedtfeld, I. (2014). Willing to give but not to forgive: Borderline personality features and cooperative behavior. *Journal of Personality Disorders, 28,* 778–795. http://dx.doi.org/10.1521/pedi_2014_28_135

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124,* 208–224. http://dx.doi.org/10.1196/annals.1440.012

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience, 3,* 119–127. http://dx.doi.org/10.1093/scan/nsn009

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27,* 813–833. http://dx.doi.org/10.1521/soco.2009.27.6.813

Tullberg, J. (2008). Trust—The importance of trustfulness versus trustworthiness. *The Journal of Socio-Economics, 37,* 2059–2071. http://dx.doi.org/10.1016/j.socec.2007.10.004

Ullmann-Margalit, E. (2004). Trust, distrust, and in between. In R. Hardin (Ed.), *Distrust* (pp. 60–82). New York, NY: Russell Sage Foundation.

Van den Bos, W., van Dijk, E., & Crone, E. A. (2012). Learning whom to trust in repeated social interactions: A developmental perspective. *Group Processes & Intergroup Relations, 15,* 243–256. http://dx.doi.org/10.1177/1368430211418698

Van den Bos, W., Westenberg, M., van Dijk, E., & Crone, E. A. (2010). Development of trust and reciprocity in adolescence. *Cognitive Development, 25,* 90–102. http://dx.doi.org/10.1016/j.cogdev.2009.07.004

Van Gelder, J.-L., & De Vries, R. E. (2012). Traits and states: Integrating personality and affect into a model of criminal decision making. *Criminology: An Interdisciplinary Journal, 50,* 637–671. http://dx.doi.org/10.1111/j.1745-9125.2012.00276.x

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77,* 337–349. http://dx.doi.org/10.1037/0022-3514.77.2.337

Van Lange, P. A. M. (2015). Generalized trust: Four lessons from genetics and culture. *Current Directions in Psychological Science, 24,* 71–76. http://dx.doi.org/10.1177/0963721414552473

Van Lange, P. A. M., & Kuhlman, D. M. (1994). Social value orientations and impressions of partner's honesty and intelligence: A test of the might versus morality effect. *Journal of Personality and Social Psychology, 67,* 126–141. http://dx.doi.org/10.1037/0022-3514.67.1.126

Van Lange, P. A. M., & Semin-Goossens, A. (1998). The boundaries of reciprocal cooperation. *European Journal of Social Psychology, 28,* 847–854. http://dx.doi.org/10.1002/(SICI)1099-0992(199809/10)28:5<847::AID-EJSP886>3.0.CO;2-L

Van Lange, P. A. M., Van Vugt, M., Meertens, R. M., & Ruiter, R. A. C. (1998). A social dilemma analysis of commuting preferences: The roles of social value orientation and trust. *Journal of Applied Social Psychology, 28,* 796–820. http://dx.doi.org/10.1111/j.1559-1816.1998.tb01732.x

Van Lange, P. A. M., Vinkhuyzen, A. A. E., & Posthuma, D. (2014). Genetic influences are virtually absent for trust. *PLoS ONE, 9,* e93880. http://dx.doi.org/10.1371/journal.pone.0093880

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108,* 796–803. http://dx.doi.org/10.1016/j.cognition.2008.07.002

Volk, S., Thöni, C., & Ruigrok, W. (2011). Personality, personal values and cooperation preferences in public goods games: A longitudinal study. *Personality and Individual Differences, 50,* 810–815. http://dx.doi.org/10.1016/j.paid.2011.01.001

Vollan, B. (2011). The difference between kinship and friendship: (Field-) experimental evidence on trust and punishment. *The Journal of Socio-Economics, 40,* 14–25. http://dx.doi.org/10.1016/j.socec.2010.10.003

Vyrastekova, J., & Garikipati, S. (2005). *Beliefs and trust: An experiment*. Discussion Paper 88, Center for Economic Research, Tilburg University. http://dx.doi.org/10.2139/ssrn.800424

Walker, D. F., & Gorsuch, R. L. (2002). Forgiveness within the Big Five personality model. *Personality and Individual Differences, 32,* 1127–1137. http://dx.doi.org/10.1016/S0191-8869(00)00185-9

Wang, C. S., Galinsky, A. D., & Murnighan, J. K. (2009). Bad drives psychological reactions, but good propels behavior: Responses to honesty and deception. *Psychological Science, 20,* 634–644. http://dx.doi.org/10.1111/j.1467-9280.2009.02344.x

Wang, T.-W. (2008). Forgiveness and Big Five personality traits among Taiwanese undergraduates. *Social Behavior and Personality, 36,* 849–850. http://dx.doi.org/10.2224/sbp.2008.36.6.849

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15,* 263–290. http://dx.doi.org/10.1002/bdm.414

Weber, J. M., Malhotra, D., & Murnighan, J. K. (2005). Normal acts of irrational trust: Motivated attributions and the trust development process. In B. M. Staw & R. M. Kramer (Eds.), *Research in organizational behavior: An annual series of analytical essays and critical reviews* (Vol. 26, pp. 75–101). Philadelphia, PA: Elsevier Science/JAI Press.

Weller, J. A., & Thulin, E. W. (2012). Do honest people take fewer risks? Personality correlates of risk-taking to achieve gains and avoid losses in HEXACO space. *Personality and Individual Differences, 53,* 923–926. http://dx.doi.org/10.1016/j.paid.2012.06.010

Weller, J. A., & Tikir, A. (2011). Predicting domain-specific risk taking with the HEXACO personality structure. *Journal of Behavioral Decision Making, 24,* 180–201. http://dx.doi.org/10.1002/bdm.677

Wilke, H. A. M. (1991). Greed, efficiency and fairness in resource management situations. *European Review of Social Psychology, 2,* 165–187. http://dx.doi.org/10.1080/14792779143000051

Wray, L. D., & Stone, E. R. (2005). The role of self-esteem and anxiety in decision making for self versus others in relationships. *Journal of Behavioral Decision Making, 18,* 125–144. http://dx.doi.org/10.1002/bdm.490

Wrightsman, L. S. J., Jr. (1964). Measurement of philosophies of human nature. *Psychological Reports, 14,* 743–751. http://dx.doi.org/10.2466/pr0.1964.14.3.743

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51,* 110–116. http://dx.doi.org/10.1037/0022-3514.51.1.110

Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly, 51,* 265–271. http://dx.doi.org/10.2307/2786924

Yamagishi, T. (2011). *Trust: The evolutionary game of mind and society*. Tokyo: Springer. http://dx.doi.org/10.1007/978-4-431-53936-0

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., & Shinada, M., . . . Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 109,* 20364–20368. http://dx.doi.org/10.1073/pnas.1212126109

Yamagishi, T., Kanazawa, S., Mashima, R., & Terai, S. (2005). Separating trust from cooperation in a dynamic relationship: Prisoner's dilemma with variable dependence. *Rationality and Society, 17,* 275–308. http://dx.doi.org/10.1177/1043463105055463

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., . . . Simunovic, D. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes, 120,* 260–271. http://dx.doi.org/10.1016/j.obhdp.2012.06.002

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion, 18,* 129–166. http://dx.doi.org/10.1007/BF02249397

Yates, J. F., & Stone, E. R. (1992). The risk construct. In J. F. Yates (Ed.), *Risk-taking behavior* (pp. 1–25). Oxford, UK: Wiley.

Yu, M., Saleem, M., & Gonzalez, C. (2014). Developing trust: First impressions and experience. *Journal of Economic Psychology, 43,* 16–29. http://dx.doi.org/10.1016/j.joep.2014.04.004

Zand, D. E. (1972). Trust and managerial problem solving. *Administrative Science Quarterly, 17,* 229–239. http://dx.doi.org/10.2307/2393957

Zettler, I., Hilbig, B. E., & Heydasch, T. (2013). Two sides of one coin: Honesty-Humility and situational factors mutually shape social dilemma decision making. *Journal of Research in Personality, 47,* 286–295. http://dx.doi.org/10.1016/j.jrp.2013.01.012

Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review, 19,* 277–302. http://dx.doi.org/10.1177/1088868314553709

Zuckerman, M., Eysenck, S., & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology, 46,* 139–149. http://dx.doi.org/10.1037/0022-006X.46.1.139

Received February 21, 2015
Revision received April 28, 2015
Accepted May 4, 2015 ■

Journal of Research in Personality 50 (2014) 61–65

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Brief Report

Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations

Isabel Thielmann *, Benjamin E. Hilbig

University of Mannheim, Mannheim, Germany

ARTICLE INFO

Article history:
Available online 22 March 2014

Keywords:
Trust
Trustworthiness expectations
Social projection
Personality
HEXACO Honesty–Humility
Cooperativeness
Economic games

ABSTRACT

Although trust is a key aspect of social behavior, individual differences in trust are not yet sufficiently understood. Addressing this issue, the present study investigated the link between trait Honesty–Humility, behavioral tendencies in economic games, and trustworthiness expectations. Based on a social projection account, it was hypothesized that individuals base their trustworthiness expectations on their own trustworthiness, i.e., their tendency to cooperate (as opposed to exploiting others). As predicted, Honesty–Humility was positively associated with trustworthiness expectations. In line with the social projection hypothesis, this relation was fully mediated through cooperativeness in the Dictator Game, but not through entitlement in the Ultimatum Game. Cooperativeness (as driven by trait Honesty–Humility) is thus an important determinant of individual differences in trust.

1. Introduction

Trust is one of the most central aspects of social behavior and has therefore been heavily studied across the social sciences and beyond. Indeed, the idea that personality plays a central role for trust was already expressed decades ago. For example, Rotter (1967) defined trust in terms of a personality trait, namely as a "general expectancy [...] that others can be relied upon" (p. 651). More commonly, though, trust has been conceptualized as a social behavior that is determined by an individual's general willingness to trust others – her so-called *trust propensity* (e.g., Mayer, Davis, & Schoorman, 1995). Despite the long history of trust-related personality research, however, the (basic) trait determinants underlying these individual differences in trust are not yet sufficiently clarified. Addressing this issue, the current work examines the underlying personality dimensions of trustworthiness expectations. These reflect individuals' beliefs about a trustee's trustworthiness and thus form a central prerequisite of trust behavior alongside the willingness to take the risk associated with trusting (e.g., Boon & Holmes, 1991). Specifically, we investigated the influence of individuals' own trait cooperativeness on trustworthiness expectations as a path of social projection.

Social projection has been discussed as a vital determinant of expectations in general, and trustworthiness expectations in particular (e.g., Krueger, Massey, & DiDonato, 2008). In situations of trust – which are defined by insufficient knowledge on others' trustworthiness – people are assumed to form corresponding expectations by projecting their own cooperativeness (or trustworthiness, respectively[1]) onto others. As such, cooperative individuals should expect others to be cooperative, and thus trustworthy, as well; uncooperative individuals, in turn, should expect others to be uncooperative, and thus untrustworthy. Correspondingly, an individual's own cooperativeness is assumed to form a basis of her expectations about a trustee's likely behavior.

In line with this idea, individuals' own cooperativeness has repeatedly been identified as a determinant of trust in strangers. For example, trustees returning large amounts to the trustor in the Trust Game (i.e., trustworthy individuals; Berg, Dickhaut, & McCabe, 1995), were found to be more willing to trust an unknown other as compared to trustees returning only small amounts or nothing (e.g., Evans & Revelle, 2008; Yamagishi et al., 2013). Similarly, a pro-social (as opposed to a pro-self) social value orientation as well as a high willingness to cooperate (in economic games) had a positive effect on individuals' willingness to trust (e.g., Kanagaretnam, Mestelman, Nainar, & Shehata, 2009; Yamagishi et al., 2013). This suggests that individual differences in cooperativeness can account for individual differences in trust.

However, previous studies did not disentangle trustworthiness expectations from trust and/or cooperative behavior as they did not assess participants' expectations about the trustee's

* Corresponding author. Address: Department of Psychology, School of Social Sciences, University of Mannheim, D7, 27, 68159 Mannheim, Germany. Fax: +49 621 181 3997.
E-mail address: isabel.thielmann@psychologie.uni-mannheim.de (I. Thielmann).

[1] In a situation of trust, a trustee's cooperativeness ultimately corresponds to her trustworthiness. Thus, we use both terms synonymously in what follows.

http://dx.doi.org/10.1016/j.jrp.2014.03.006
0092-6566/

trustworthiness prior to the decision whether to trust or not. For example, a large investment in the Trust Game may either indicate an optimistic expectation about the trustee's likely return or, instead, a high willingness to take the risk associated with unreciprocated trust. Likewise, large investments may be driven by a high willingness to share – simply as an expression of individuals' cooperativeness. Therefore, it remains unclear whether the observed relation between cooperativeness and trust is actually due to social projection or whether game-based cooperation and trust behavior merely share a common core in terms of trait cooperativeness.

In any case, a social projection mechanism implies that basic personality traits driving cooperation should also relate to individual differences in trust. Supporting this notion, Big Five-Agreeableness – capturing individual differences in the motivation to cooperate (Denissen & Penke, 2008) – has been identified as the main predictor of trust in the Trust Game (e.g., Evans & Revelle, 2008). However, as Big Five-Agreeableness specifically includes a *trust* facet, the mechanism underlying this relation remains inconclusive. On the one hand, it is possible that Agreeableness simply includes the tendency to trust, thus leading to more trust behavior (without any social projection involved). On the other hand, agreeable individuals should be more likely to cooperate (Denissen & Penke, 2008) and may project this tendency onto strangers, thus reflecting a social projection mechanism. In essence, the link between Big Five-Agreeableness and trust behavior cannot provide strong evidence for the hypothesis of social projection (of trait cooperativeness) as an underlying determinant of trustworthiness expectations.

A more conclusive test of social projection thus requires consideration of a trait that specifically signals cooperativeness without aspects of trust propensity. One corresponding basic trait is Honesty–Humility, the sixth dimension of the HEXACO model of personality structure (Honesty–Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to Experience; Ashton & Lee, 2007). Honesty–Humility particularly refers to an individual's cooperativeness in terms of sincerity, fairness, greed-avoidance, and modesty. As such, Honesty–Humility shares some content with Big Five-Agreeableness (i.e., sincerity and modesty), but also comprises more unique aspects (i.e., fairness and greed-avoidance) which are not captured by the Big Five factors (Ashton, Lee, & de Vries, in press).[2] In line with this conceptualization, Honesty–Humility has repeatedly and consistently been identified as a predictor of *active cooperation* in terms of fairness and non-exploitation in economic games (e.g., Hilbig, Glöckner, & Zettler, in press). Investigating the influence of Honesty–Humility on trustworthiness expectations can hence offer insights into the role of *pure* trait cooperativeness (and thus social projection) for individual differences in trust.

Based on this reasoning, the present study investigated the link between Honesty–Humility, behavioral tendencies in economic games, and trustworthiness expectations. As implied by the idea of social projection, we hypothesized Honesty–Humility to positively relate to trustworthiness expectations (Hypothesis 1). To test still more conclusively whether indeed cooperativeness links Honesty–Humility and trustworthiness expectations, we additionally considered the allocation in the Dictator Game as a measure of cooperativeness. In the Dictator Game, individuals simply divide an endowment between themselves and another person. The mechanism of social projection clearly predicts that the positive relation between Honesty–Humility and trustworthiness expectations must be mediated through cooperativeness in this game (Hypothesis 2).

Furthermore, it is necessary to rule out that entitlement (rather than cooperativeness) drives said relation between Honesty–Humility and trustworthiness expectations. That is, individuals high in Honesty–Humility should expect others to be trustworthy because they themselves are and *not* because they feel entitled to a good treatment by others. To rule out this mechanism, a measure of entitlement was obtained via the Ultimatum Game in which individuals indicate how much an unknown other must offer (from an initial endowment) for them to accept the offer. Prior research has already indicated that Honesty–Humility and Ultimatum Game acceptance levels are indeed unrelated (e.g., Hilbig, Zettler, Leist, & Heydasch, 2013). In any case, the hypothesis that social projection of cooperativeness and not entitlement links Honesty–Humility to trustworthiness expectations predicts that Ultimatum Game acceptance levels should not mediate the relationship between these two variables (Hypothesis 3).

2. Method

2.1. Participants

Participants were recruited online via internet communities and mailing lists of the University of Mannheim, Germany. An a-priori power analysis revealed that to uncover a small to medium-sized effect ($r = .20$) with optimal statistical power ($1 - \beta = .95$), a sample of about $N = 260$ was required. Expecting a typical drop-out-rate of about 15%, we recruited 301 participants. Out of these, 81% ($N = 244$) fulfilled the criteria for inclusion (i.e., completion of all tasks, no repeated participation, and at least a "good" grasp of the German language).[3] The final sample comprised 79% females, aged 18–75 years ($M = 28.79$, $SD = 10.64$). Most participants were students (50%) or employees (41%).

2.2. Materials

Basic personality traits (including Honesty–Humility) were measured via the German 60-item version (Moshagen, Hilbig, & Zettler, in press) of the HEXACO Personality Inventory-Revised (HEXACO-60; Ashton & Lee, 2009). The inventory contains 10 items for each of the six HEXACO dimensions. To assess trustworthiness expectations, cooperativeness, and entitlement, we used three different games: the Distrust Game, the Dictator Game, and the Ultimatum Game. All games were completely hypothetical. That is, participants were asked to imagine playing each game with another unknown person for money.

In the Distrust Game (McEvily, Radzevick, & Weber, 2012), two players (trustor and trustee) each receive an initial endowment of, say, 50€. However, the trustee is empowered to take any amount of the 50€ initially assigned to the trustor, in turn increasing her own payoff by decreasing the trustor's payoff. For example, if the trustee decides to take 30€ from the trustor, she receives 80€ in total whereas the trustor ends up with 20€. Hence, the trustor's payoff depends on the trustee's trustworthiness in terms of her willingness to maintain the fair split as opposed to taking some of the trustor's endowment. This was thoroughly explained to participants. As a measure of trustworthiness expectations, participants

[2] As the similar names suggest, there is also substantial overlap between Big Five- and HEXACO-Agreeableness. However, the two are not equivalent: Whereas they share content such as forgiveness and gentleness, HEXACO-Agreeableness also covers even-temper versus irritability, anger, and harshness – which is considered to belong to Neuroticism in the Big Five. In turn, sentimentality-related content, which is associated with Big Five-Agreeableness, is not captured by HEXACO-Agreeableness but instead included in the Emotionality factor of the HEXACO model (i.e., the counterpart of Big Five-Neuroticism; Ashton et al., in press).

[3] Note that although this sample size is slightly below the optimum determined through the power analysis, it nonetheless yields a highly satisfactory power of $1-\beta = .93$.

(trustors) were asked to estimate how much money they expected a hypothetical random trustee to leave for them – or, in other words, how much the trustee would not take from them for her personal profit. Participants indicating that the trustee will take away nothing (thus leaving the equal split untouched) are considered to have high trustworthiness expectations. In turn, if expecting an untrustworthy interaction partner, participants should indicate that the trustee will take some of their endowment, thus leaving less than 50€. So, the higher the amount a trustor expected the trustee to leave, the higher the trust in the trustee's trustworthiness. As a plausibility check (and in line with the original version of the game), we further asked those participants who expected the trustee to behave untrustworthily how much of their own endowment they would be willing to give up for their protection (thus eliminating the trustee's power over the final distribution). In general, the Distrust Game offers straightforward assessment of trustworthiness expectations because the trustee's behavior is not contingent on the trustor's behavior.

In addition to the Distrust Game, we used a hypothetical Dictator Game (Forsythe, Horowitz, Savin, & Sefton, 1994) as a measure of cooperativeness. Participants played the allocator who is asked to divide an amount of 100€ between herself and a hypothetical recipient. As an advantage over other games (e.g., Prisoner's Dilemma Game), the allocation in the Dictator Game is a pure measure of cooperativeness that is not confounded with trust.

Finally, to rule out entitlement as an alternative explanation (cf. Hypothesis 3), we collected participants' responses in a hypothetical Ultimatum Game (Güth, Schmittberger, & Schwarze, 1982). The recipient in the Ultimatum Game is empowered to either accept or reject a proposer's offer on how to split an initial endowment between the two. Whereas accepting means that the offer is realized as proposed, rejecting means that both players receive nothing. The recipient can hence punish the proposer for an offer that does not satisfy her demands. Again, this was thoroughly explained to participants and they were asked to indicate the minimum offer they would be willing to accept, implying that they would reject any offer below this minimum. The higher the offer one is minimally willing to accept, the higher one's subjective entitlement.

2.3. Procedure

The study was run as a web-based study in close adherence to the proposed standards for web-based experimenting (Reips, 2002). After providing informed consent and demographical information, participants completed the HEXACO-60 as well as other personality measures not pertinent to this investigation. Following the questionnaires, participants worked on the Distrust Game, the Dictator Game, and the Ultimatum Game, respectively. After the games, participants answered eight control questions as a measure of their seriousness of participation. Finally, as an incentive for participation, they received feedback on their HEXACO personality profile. On average, completion of the entire study took 21 min.

3. Results

Taking into account potential measurement error in the observed variables, we resorted to an exploratory structural equation modeling (ESEM) approach[4] with latent factors for the six HEXACO dimensions. Considering that personality dimensions are not perfectly orthogonal and therefore usually correlated (e.g., Ashton, Lee, Goldberg, & de Vries, 2009), ESEM allows cross-loadings of items on other than their primary factor. As a consequence, ESEM solves part of the typical problems of confirmatory factor analyses, such as poor model fit, inflated factor correlations, and biased parameter estimates (Asparouhov & Muthén, 2009; see also for more information on ESEM). ESEM was run with an oblique target rotation method, setting target values for all items except the ones intended to load on the respective HEXACO factor to zero. Overall, the model yielded a satisfactory fit as all fit indices recommended for personality research (Beauducel & Wittmann, 2005) were in an acceptable range, $\chi^2/df = 1.80$, RMSEA = .06, and SRMR = .05. Supporting the factor structure of the HEXACO-60 and the interpretation of the latent factors, each item had its primary loading on the corresponding latent trait dimension.

Table 1 provides the means, standard deviations, and correlations for all measures of interest (for correlations between all variables see Table S1 in the online supplemental material). In general, participants were rather optimistic about the hypothetical other's trustworthiness in the Distrust Game. That is, about two thirds of participants ($n = 167$) expected the trustee to leave the fair split untouched whereas only one third of participants ($n = 77$) expected the trustee to take away some of their endowment, namely on average 50.6% ($SD = 27.6\%$) of it. Confirming that participants understood the mechanism of the game, the amount individuals expected the trustee to take away strongly predicted the amount they were willing to pay for their protection, $r = .49$, $p < .001$.

As can be seen in Table 1, the pattern of results was in line with the hypotheses: With regard to Hypothesis 1, we found a small to medium-sized positive relation ($r = .20$, 95% CI [.07,.34]) between Honesty–Humility and trustworthiness expectations, showing that individuals high in Honesty–Humility expected the trustee to leave a higher amount for them than individuals low in Honesty–Humility. The effect of Honesty–Humility on trustworthiness expectations remained significant when controlling for the remaining HEXACO dimensions as covariates, $\beta = .16$, $p = .033$. Otherwise, only Extraversion explained unique variance in trustworthiness expectations, $\beta = .14$, $p = .039$.

In line with Hypothesis 2, we found positive associations between Honesty–Humility and cooperativeness (i.e., Dictator Game allocation) and between cooperativeness and trustworthiness expectations (Table 1). As predicted by the proposed social projection mechanism, a mediation analysis revealed a significant indirect effect from Honesty–Humility to trustworthiness expectations via Dictator Game cooperativeness, $a^* b = 0.16$, 95% CI [0.08, 0.23], $p < .001$.[5] In terms of the widely-accepted causal steps approach for testing mediation (Baron & Kenny, 1986), all criteria for a *full mediation* were fulfilled (cf. Fig. 1).

Finally, Hypothesis 3 was also corroborated as neither a relation between Honesty–Humility and entitlement (i.e., Ultimatum Game acceptance level) nor between entitlement and trustworthiness expectations was found (Table 1).[6] Correspondingly, a mediation analysis yielded a non-significant indirect effect from Honesty–Humility to trustworthiness expectations via Ultimatum Game entitlement, $a^* b = 0.01$, 95% CI [−0.01, 0.02], $p = .395$ (cf. Fig. 1).

4. Discussion

Although trust has been heavily studied across various disciplines, individual differences in trust are insufficiently understood.

[4] Note that all results reported in the following were replicated with traditional scale-based analyses.

[5] The indirect effect refers to the product of regression coefficients when regressing the mediator on the predictor (a) and the criterion on the mediator (b). Repeating mediation analyses with bootstrapped confidence intervals for indirect effects in a traditional SEM framework confirmed the obtained result.

[6] A Bayesian analysis (Wagenmakers, 2007) clearly favored the null hypothesis over the alternative hypothesis for both correlations ($p(H_0|D) = 72\%$ and $p(H_0|D) = 85\%$, respectively).

Table 1
Means, standard deviations (in parentheses), and intercorrelations between all measures.

Measure	Scale	*M* (*SD*)	Correlations			
			1	2	3	4
1. Honesty–Humility[a]	1–5	3.54 (0.61)	–			
2. Distrust Game expectation	0–50	7.98 (14.09)	.20**	–		
3. Dictator Game cooperativeness	0–100	42.19 (15.04)	.43***	.36***	–	
4. Ultimatum Game entitlement	0–100	29.15 (17.60)	.12	.09	.08	–

** $p < .01$.
*** $p < .001$.
[a] Cronbach's alpha for Honesty–Humility was .75.

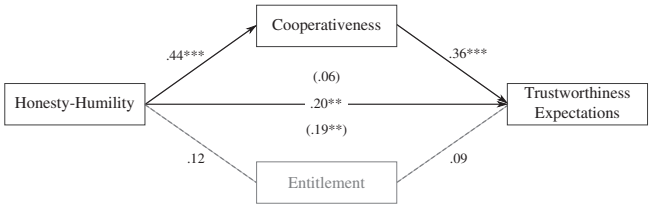

Fig. 1. Mediation model of Honesty–Humility to trustworthiness expectations with corresponding ESEM-based correlation and partial correlation coefficients (in parentheses), controlling for the respective mediator variable. Solid connections indicate significant relations; dashed connections indicate null relations (as confirmed by Bayesian analyses).

Addressing this issue, the present study investigated the link between the HEXACO personality traits, cooperativeness versus entitlement, and trustworthiness expectations – a key aspect of trust (e.g., Boon & Holmes, 1991). As derived from the idea of social projection, it was hypothesized that individuals' tendency to cooperate (as driven by trait Honesty–Humility) should explain individual differences in expectations about strangers' trustworthiness.

In line with hypotheses – and mirroring a path of social projection – individuals' trustworthiness expectations were related to their own cooperativeness which, in turn, was related to trait Honesty–Humility. That is, individuals high in Honesty–Humility expected an unknown trustee to be more trustworthy than individuals low in Honesty-Humility. This relation was fully mediated through cooperativeness (as measured via the Dictator Game), but not through entitlement (as measured via the Ultimatum Game). Hence, individuals high in Honesty–Humility tended to express more optimistic trustworthiness expectations – not because they thought that others *should* but rather because they themselves *would* behave cooperatively. Stated differently, their trustworthiness expectations were well accounted for by their own cooperativeness and clearly not based on a feeling of entitlement, i.e., on how much cooperation they thought they deserved from another. Overall, the pattern supports the notion that social projection is indeed a relevant mechanism underlying the relation between cooperativeness (as predicted by trait Honesty–Humility) and trustworthiness expectations. Besides, it nicely mirrors the theoretical conceptualization of Honesty–Humility capturing *fairness* on the one hand and *modesty* on the other hand.

As a limitation, it must be acknowledged that correlation-based mediation analyses cannot confirm any form or direction of causation (cf. MacKinnon, Fairchild, & Fritz, 2007). Nonetheless, the main hypothesis of this work strictly predicts a certain path (and thus a specific mediation pattern). In turn, absence of the mediation pattern would have spoken directly against the hypothesis and thus it represents an appropriate test, despite the inherent limitations in terms of causal inference. Similarly, we cannot fully rule out that individuals based their self-reports on their perceptions and/or expectations of others rather than – as implied by the idea of social projection – basing their expectations of others on what they themselves would do. However, given that individuals could not observe any actions of another and were entirely unaware of who the other person might be, it seems unlikely that specific perceptions or expectations of others can account for the relation between self-reported trait cooperativeness and trustworthiness expectations.

Furthermore, it is important to take into account that individuals may differ in their tendency to project own characteristics onto others (Krueger & Acevedo, 2007). Therefore, future research might more conclusively test the role of social projection by considering such differences in *social projection tendency* as a potential moderator of the relation between traits driving cooperativeness and trustworthiness expectations. The potentially moderating role of social projection tendency might also explain why the effect of Honesty–Humility on trustworthiness expectations was only small to medium in size.

Beyond the current hypotheses, we also found a positive effect of Extraversion on trustworthiness expectations. As high levels of Extraversion are associated with feelings of confidence and optimism – and therefore arguably with trust (Ashton et al., in press) – this finding is theoretically plausible and corroborates the usefulness of the HEXACO model for studying individual differences in trust at a high resolution. Nonetheless, it would seem worthwhile for future studies to dissect such nuances more directly, e.g., by comparing the HEXACO and Big Five models in terms of which specific aspects of personality explain individual differences in trust (cf. Hilbig et al., in press).

In conclusion, the present study identified trait cooperativeness (i.e., Honesty–Humility) as a basic personality dimension underlying trustworthiness expectations. Thus, the HEXACO model also seems to provide useful and novel insights concerning individual differences in trust especially due to the inclusion of a sixth basic factor specifically capturing pro-social tendencies (e.g., Hilbig et al., in press). Given the vital importance of trust for all kinds of social interactions, it is our hope that the current approach encourages future research on individual differences in trust in general, also beyond the current focus on trustworthiness expectations.

Acknowledgments

The work reported herein was supported by Grants to the second author from the Baden-Württemberg Foundation (Germany) and the German Research Foundation [HI 1600/1-1].

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jrp.2014.03.006.

References

Ashton, M. C., Lee, K., & de Vries, R. E. (in press). The HEXACO Honesty–Humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*(4), 340–345.

Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher order factors of personality: Do they exist? *Personality and Social Psychology Review, 13*(2), 79–91.

Asparouhov, T., & Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural equation modeling, 16*(3), 397–438.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*(1), 41–75.

Berg, J., Dickhaut, J., & McCabe, K. A. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142.

Boon, S. D., & Holmes, J. G. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In R. A. Hinde & J. Groebel (Eds.), *Cooperation and prosocial behavior* (pp. 190–211). Cambridge, England: Cambridge University Press.

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor Model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality, 42*(5), 1285–1302.

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality, 42*(6), 1585–1593.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*(3), 347–369.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*(4), 367–388.

Hilbig, B. E., Glöckner, A., & Zettler, I. (in press). Personality and pro-social behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology*.

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences, 54*(5), 598–603.

Kanagaretnam, K., Mestelman, S., Nainar, K., & Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. *Journal of Economic Psychology, 30*(3), 368–380.

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning. *The American Journal of Psychology, 120*(4), 593–618.

Krueger, J. I., Massey, A. L., & DiDonato, T. E. (2008). A matter of trust: From social preferences to the strategic adherence to social norms. *Negotiation and Conflict Management Research, 1*(1), 31–52.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593–614.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709–734.

McEvily, B., Radzevick, J. R., & Weber, R. A. (2012). Whom do you distrust and how much does it cost? An experiment on the measurement of trust. *Games and Economic Behavior, 74*(1), 285–298.

Moshagen, M., Hilbig, B. E., & Zettler, I. (in press). Faktorstruktur, psychometrische Eigenschaften und Messinvarianz der deutschen Version des 60-Item HEXACO Persönlichkeitsinventars. *Diagnostica*.

Reips, U.-D. (2002). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review, 20*(3), 241–249.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*(4), 651–665.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779–804.

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., et al. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes, 120*(2), 260–271.

Article

The Traits One Can Trust: Dissecting Reciprocity and Kindness as Determinants of Trustworthy Behavior

Personality and Social Psychology Bulletin
2015, Vol. 41(11) 1523–1536

DOI: 10.1177/0146167215600530
pspb.sagepub.com
SAGE

Isabel Thielmann[1,2] and Benjamin E. Hilbig[1]

Abstract

Trustworthiness is a vital pillar of various social interactions hinging upon trust. However, the underlying determinants of trustworthiness—especially in terms of (basic) personality traits—are insufficiently understood. Specifically, three mechanisms underlying trustworthiness have been proposed: unconditional kindness, positive reciprocity, and negative reciprocity. The present research aims to disentangle these mechanisms using a trait-based approach, relying on the HEXACO (Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to Experience) model of personality. In three studies, participants acted as the trustee in the trust game. All studies revealed consistent support for the unconditional kindness mechanism, showing an exclusive link between Honesty-Humility and trustworthiness, irrespective of the level of prior trust. In turn, positive and negative reciprocity could not account for the pattern of results. In addition, our results reconcile the inconsistent evidence on the relation between Big Five-Agreeableness and trustworthiness: Unconditional kindness only refers to one component of the broad Big Five-Agreeableness factor (which subsumes various cooperative tendencies).

Keywords

trustworthiness, trust game, reciprocity, unconditional kindness, basic personality traits

Received March 6, 2015; revision accepted July 22, 2015

Trust is a vital pillar of various social interactions and society as a whole (e.g., Yamagishi, 2011). However, the benefits associated with trust are necessarily contingent upon others' trustworthiness—given that trust basically mirrors a risky choice to depend on another without being able to control the other's actions (Thielmann & Hilbig, in press). That is, trust—especially among strangers—is only defensible if one can expect the trusted person (the so-called *trustee*) to honor rather than betray the trust (Evans & Krueger, 2009). Despite this vital significance of trustworthiness, there is only a rudimentary understanding of trustworthy behavior, especially in terms of underlying personality traits.

Inspired by behavioral economics, recent research on trustworthiness has mostly relied on the *trust game* (Berg, Dickhaut, & McCabe, 1995). In this game, a trustor is asked to divide a certain endowment between herself and a trustee. The amount the trustor en*trusts* is multiplied (usually tripled) and transferred to the trustee who is then asked to decide how much to return to the trustor. By implication, the amount returned is considered a measure of behavioral trustworthiness, with high returns indicating high trustworthiness (cf. Johnson & Mislin, 2011). Note that, according to this conceptualization, we herein adopt a behavioral view of trust and trustworthiness.

As directly follows from the rules of the trust game, trustworthiness involves a *reaction*, that is, behavior contingent on another's (prior) trust. Correspondingly, trustworthiness in the trust game has typically been considered an expression of reciprocity (Berg et al., 1995), which can be defined as "a conditional behaviour aimed at reacting to a behaviour with another behaviour of the same valence" (Perugini & Gallucci, 2001, p. S20). Stated differently, reciprocity captures an individual's tendency to adjust her own behavior to an interaction partner's (previous) behavior. Depending on whether an individual rewards another's cooperative behavior or punishes another's uncooperative behavior, one can further distinguish between positive and negative reciprocity (e.g., Perugini, Gallucci, Presaghi, & Ercolani, 2003). Although recent research on the trust game almost exclusively interpreted trustworthiness in terms of positive reciprocity (e.g., Chaudhuri & Gangadharan, 2007; McCabe, Rigdon, & Smith, 2003), we consider both

[1]University of Koblenz-Landau, Germany
[2]University of Mannheim, Germany

Corresponding Author:
Isabel Thielmann, University of Koblenz-Landau, Fortstraße 7, 76829 Landau, Germany.
Email: thielmann@uni-landau.de

positive and negative reciprocity as potential factors accounting for trustworthy behavior.[1]

The idea that trustworthiness essentially mirrors reciprocity is largely based on evidence indicating that higher trust levels enhance trustees' willingness to behave trustworthily (see Johnson & Mislin, 2011, for a meta-analytic review). Likewise, if trustees highly (rather than only marginally) benefitted from the trustor's trust, trustworthiness increased (Malhotra, 2004). Vice versa, if trustees cannot rule out that trustors merely "trusted" out of strategic considerations rather than out of "true" trust or kindness, respectively, returns have been shown to decrease (Bauernschuster, Falck, & Große, 2013). Finally, trustee returns have been negatively related to trait negative reciprocity; however, for positive reciprocity a comparable (positive) link could not be corroborated (Yamagishi et al., 2012). Nevertheless, altogether, these findings suggest a mechanism of reciprocity underlying trustworthiness—implying that dispositions toward (positive or negative) reciprocity should determine individual differences in trustworthy behavior.

Besides reciprocity, it has also been argued—and empirically supported—that trustee returns are driven by individuals' unconditional kindness (e.g., Cox, 2004; Gambetta & Przepiorka, 2014), which has mostly been operationalized through giving in the dictator game[2] (Forsythe, Horowitz, Savin, & Sefton, 1994). Unconditional kindness implies that trustworthy behaviors are not perfectly contingent on the level of prior trust, but rather involve a relatively stable return. Correspondingly, meta-analytic evidence indicates that trustee returns decline less than proportionately with the multiplier of the entrusted amount (Johnson & Mislin, 2011). This implies relatively high returns if the amount is, for example, only doubled rather than tripled—a finding that is difficult to explain by reciprocity alone. Similarly, directly comparing (game-based) reciprocity and unconditional kindness revealed that the latter accounts for the majority of variance in trustworthiness (Ashraf, Bohnet, & Piankov, 2006). However, in another study, neither unconditional kindness nor reciprocity showed significant relations with trustworthiness (Ben-Ner & Halldorsson, 2010). In any case, an unconditional kindness mechanism would imply that dispositions toward altruistic and fair behavior can explain individual variation in trustworthiness.

Taken together, the extant evidence is inconclusive regarding the nature and underlying (trait) determinants of trustworthiness. Strikingly, though, almost all previous studies have exclusively relied on a game-theoretical approach, for example, by investigating behavioral tendencies across different games (or structural changes within one game). This common practice in behavioral economics is undoubtedly fruitful, but essentially misses out on more stable behavioral tendencies (i.e., traits) and may additionally yield caveats due to common-method variance and individuals' desire to respond consistently across games. Besides, given that a zero return in the trust game might reflect either negative reciprocity or a lack of positive reciprocity, distinguishing between positive and negative reciprocity is impossible using a purely game-based approach unless the game structure is changed considerably—which would, in turn, undermine comparability across studies.

As an alternative to the sole reliance on games, models of basic personality traits offer a promising avenue to identify the determinants underlying cooperation in general and trustworthiness in particular. Specifically, "broad and stable interpersonal traits can help explain behavioral heterogeneity across a range of games modeling social interactions" (Zhao & Smillie, 2015, p. 293). Regarding trustworthiness, most corresponding research focused on the widely accepted Five-Factor Model (FFM; Costa & McCrae, 1992; McCrae & Costa, 1985) and linked trustee behavior to Agreeableness (FFM-AG; see Zhao & Smillie, 2015, for a recent review). By definition, FFM-AG captures the tendency to cooperate in situations involving resource conflicts (Denissen & Penke, 2008). Correspondingly, some studies report a positive relation between FFM-AG and trustee returns (Becker, Deckers, Dohmen, Falk, & Kosse, 2012; Ben-Ner & Halldorsson, 2010). However, a similar number of studies could not corroborate said link (Evans & Revelle, 2008; Müller & Schwieren, 2012) or found that FFM-AG is only predictive in combination with other factors (low Neuroticism; Lönnqvist, Verkasalo, Wichardt, & Walkowitz, 2012). Summarized carefully, the evidence is currently inconclusive.

Plausibly, the inconsistent findings may not be due to a conceptual limitation of FFM-AG per se, but may be due to the broad nature of this factor capturing all kinds of cooperative tendencies (including unconditional kindness and positive/negative reciprocity; cf. Costa, McCrae, & Dye, 1991). Thus, the inconsistent evidence might actually suggest that only one of the proposed determinants explains trustworthy behavior. If, for example, only unconditional kindness accounts for trustworthiness whereas positive and negative reciprocity do not, the former mechanism would strengthen the association between FFM-AG and trustworthiness, whereas the latter would reduce it—leading to an inconsistent overall picture such as the one observed. In consequence, because FFM-AG covers all trait aspects potentially—but not necessarily—relevant for trustworthiness, a positive relation between FFM-AG and trustworthiness cannot help unravel which specific dispositional tendency is actually decisive.

Fortunately, the more recently proposed HEXACO (Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness to Experience) model of personality (Ashton & Lee, 2007; Lee & Ashton, 2004) distinguishes between three trait dimensions accounting for individual variation in prosocial behavior: Honesty-Humility, Emotionality, and Agreeableness (Ashton, Lee, & De Vries, 2014). Whereas Emotionality involves a tendency toward kin altruism, Honesty-Humility (HEX-HH) and Agreeableness (HEX-AG) encompass complementary aspects of reciprocal

altruism. That is, HEX-HH is defined as "the tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation" (Ashton & Lee, 2007, p. 156). As such, high levels of HEX-HH imply sincerity, fairness, greed-avoidance, and modesty. HEX-AG, in turn, describes "the tendency to be forgiving and tolerant of others, in the sense of cooperating with others even when one might be suffering exploitation by them" (Ashton & Lee, 2007, p. 156). Thus, high levels of HEX-AG stand for forgiveness, gentleness, flexibility, and patience.

Corroborating the theoretical conceptualizations of both HEXACO factors with regard to prosocial behavior, HEX-HH has consistently (positively) been linked to active cooperation (i.e., non-exploitation) in social dilemmas (Hilbig, Zettler, & Heydasch, 2012; Zettler, Hilbig, & Heydasch, 2013) and—of particular interest for the issue at hand—to unconditional kindness in the dictator game (e.g., Baumert, Schlösser, & Schmitt, 2013; Hilbig, Thielmann, Hepp, Klein, & Zettler, 2015; Thielmann & Hilbig, 2014) as well as to positive reciprocity (Ackermann, Fleiß, & Murphy, in press; Perugini et al., 2003). For HEX-AG, in turn, studies point to negative associations with reactive cooperativeness (i.e., non-retaliation) in the ultimatum game (Hilbig, Zettler, Leist, & Heydasch, 2013; Thielmann, Hilbig, & Niedtfeld, 2014) and, specifically, negative reciprocity (Ackermann et al., in press; Perugini et al., 2003).

Due to this specific distinction between different cooperative tendencies captured by HEX-HH and HEX-AG, respectively, the HEXACO model allows for a particularly fine-grained analysis of dispositional tendencies underlying cooperative behavior. The model is thus especially useful whenever evidence based on the broader FFM is inconclusive (e.g., Hilbig, Glöckner, & Zettler, 2014). Correspondingly, for trustworthiness, specific predictions on the relation between trustworthy behavior and HEX-HH or HEX-AG, respectively, can be derived for each of the proposed potential mechanisms (as detailed below). Exactly this type of evidence—on the link between the HEXACO dimensions and trustworthiness (in the trust game)—is currently missing (Zhao & Smillie, 2015).

Consequently, the purpose of the present studies was to dissect the potential dispositional tendencies underlying trustworthiness based on the HEXACO model of personality. Given the theoretical conceptualizations of HEX-HH and HEX-AG—and corresponding evidence as sketched above—the following predictions can be derived: If trustworthiness is determined by unconditional kindness (alone), it should be positively linked to HEX-HH, but *not* linked to HEX-AG (as the latter refers to reactive cooperativeness which is, by definition, conditional). A similar main effect of HEX-HH would be compatible—but not necessary—if positive reciprocity is the responsible factor. Nonetheless, the unconditional kindness and positive reciprocity mechanisms make incompatible predictions on the presence of an interaction with the level of prior trust: By definition, unconditional kindness is unconditional and it should therefore drive trustworthiness independently of the trustor's level of trust (precluding an interaction). By contrast, positive reciprocity is inherently conditional and should thus drive trustworthiness contingent on the trustor's prior behavior (implying an interaction): The more is entrusted, the more strongly HEX-HH would have to predict trustworthiness. In summary, unconditional kindness would thus require a main effect of HEX-HH on trustworthiness, but none of HEX-AG, and no interaction of HEX-HH with prior trust. Positive reciprocity, in turn, would require said interaction; otherwise, a main effect of HEX-HH would be compatible, but not necessary.

Finally, if trustworthiness is determined by negative reciprocity, it should be (negatively) linked to HEX-AG. However, this relationship must also be a conditional one (given the conditional nature of reciprocity), implying an interaction between HEX-AG and prior trust (i.e., a stronger relation between HEX-AG and trustworthiness with decreasing levels of trust).[3] In turn, a main effect of HEX-AG on trustworthiness would not be required, but nonetheless be compatible with the negative reciprocity account.

To test the alternative mechanisms, we conducted three online studies on the link between basic personality traits and trustworthiness—all in close adherence to standards for Internet-based experimenting (e.g., Reips, 2002). As our primary goal was to disentangle the different potential determinants of trustworthy behavior, we exclusively focused on the relations between the HEXACO dimensions (particularly, HEX-HH and HEX-AG) and trustworthiness in Study 1. In Studies 2 and 3, we additionally aimed at clarifying whether the relatively weak link between FFM-AG and trustworthiness (Zhao & Smillie, 2015) can be attributed to the broader nature of FFM-AG (subsuming different cooperative tendencies).

Study 1

Method

Materials. To assess basic personality traits, we used the German 60-item version of the HEXACO Personality Inventory–Revised (HEXACO-60; Ashton & Lee, 2009; for psychometric properties of the German version, see Moshagen, Hilbig, & Zettler, 2014). The HEXACO-60 includes 10 items for each of the six HEXACO dimensions. Responses are given on a 5-point Likert-type scale ranging from *strongly disagree* to *strongly agree*.

To measure trustworthiness, we relied on the classical version of the trust game (Berg et al., 1995), placing participants in the role of the trustee.[4] During the game, participants could earn points, based on which they were later incentivized. Specifically, each participant was randomly matched to an unknown trustor. The combination of the participant's and the trustor's choices determined participants' point scores.

Table 1. Means, Standard Deviations (in Parentheses), and Bivariate Correlations (95% Confidence Intervals in Brackets) of All Focal Variables Assessed in Study 1, With Internal Consistency Reliabilities (Cronbach's α) in the Diagonal.

			Correlations	
Measure	Scale	*M* (*SD*)	HEX-HH	HEX-AG
HEX-HH	1-5	3.43 (0.64)	.80	
HEX-AG	1-5	3.08 (0.59)	.24** [.06, .41]	.78
Return 15 points (in %)	0-100	38.9 (23.6)	.35*** [.18, .51]	.00 [−.19, .19]
Return 30 points (in %)	0-100	40.3 (23.3)	.42*** [.26, .57]	.06 [−.13, .25]
Return 45 points (in %)	0-100	41.2 (21.1)	.39*** [.22, .54]	.08 [−.11, .26]
Return 60 points (in %)	0-100	42.8 (21.1)	.40*** [.22, .54]	.07 [−.12, .26]
Return 75 points (in %)	0-100	42.8 (22.0)	.34*** [.16, .50]	.05 [−.14, .24]
Return 90 points (in %)	0-100	43.8 (23.5)	.39*** [.22, .54]	.06 [−.13, .24]
M return (in %)	0-100	41.6 (20.5)	.42*** [.25, .56]	.06 [−.13, .24]

Note. HEX-HH = Honesty-Humility; HEX-AG = HEXACO-Agreeableness.
**$p \leq .01$. ***$p \leq .001$.

The 25% of participants with the highest final score received a 10€ (approximately US$12.60) gift voucher.

Initially, both trustors and trustees received an endowment of 30 points. Trustees (participants) were informed that the trustor (a randomly assigned unknown other, denoted as Player 1) could decide how much of this endowment (in 5-point increments) she wants to transfer to the trustee (denoted as Player 2). The transfer was tripled accordingly. Trustees' task was to indicate how much of the (tripled) transfer they wanted to return to the trustor. Corresponding to the widely accepted strategy method (Selten, 1967), participants were unaware of the trustor's actual transfer, but specified their return for each of the six potential (tripled) amounts (above 0) the trustor could transfer (i.e., between 15 and 90 points, in 15-point increments).

Procedure. After providing informed consent and demographic information, participants completed the HEXACO-60. Next, they were thoroughly introduced to the rules of the trust game and asked to indicate their return to the trustor for each potential transfer. Finally, participants answered a few control questions assessing their seriousness of participation and received individual feedback on their HEXACO scores. After completing data collection, each participant was randomly assigned to a trustor to determine their point scores (using pseudonymous codes preserving anonymity).

Participants. Participants were recruited via online social networks (e.g., Facebook) and university mailing lists. An a priori power analysis using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) revealed a required sample size of $N = 108$ to detect a medium-sized effect ($f^2 = .10$) in a two-tailed t test for single regression coefficients in a linear regression with a high power ($1 - \beta = .90$). Note that sample size calculations were based on the main effects of HEX-HH and HEX-AG, respectively, on trustworthiness, given that these mirror the most basic tests in our analyses. Corresponding to these calculations, we recruited $N = 108$ trustees (and the same number of trustors, see above), including 81 females and aged between 19 and 49 years ($M = 25.5$, $SD = 5.6$). The majority (72.2%) of participants were students, 17.6% were employees.

Results and Discussion

Table 1 summarizes the means, standard deviations, and zero-order correlations between all focal variables (for information on all variables, see Table S1 in the Supplemental Material). First off, trustee returns showed strong positive correlations with HEX-HH for all potential trust levels individually as well as for the average return across trust levels (all rs > .30, $p < .001$). For HEX-AG, in turn, no (or only very weak) correlations with trustworthiness could be observed (all rs < .08, $p > .43$).

To statistically test this pattern (i.e., an influence of HEX-HH, but not of HEX-AG, on trustworthiness), we analyzed the average return across trust levels using a three-step analytical approach (cf. Hilbig et al., 2014). First, we used an approximation of the Bayesian Information Criterion (BIC) from R^2 (Raftery, 1995, Equation 26; Wagenmakers, 2007). From the BIC, we calculated Bayes Factors (BF). Following Wagenmakers (2007), we refer to BF_{01}, relating the probability of the null hypothesis to the probability of the alternative hypothesis. Thus, $BF_{01} < 1$ indicates evidence in favor of the alternative hypothesis, whereas $BF_{01} > 1$ indicates evidence in favor of the null hypothesis. For HEX-HH, a $BF_{01} = 0.0003$ indicated that the alternative hypothesis (a meaningful correlation between HEX-HH and trustworthiness) was more than 3,000 times as likely as the null hypothesis given the data. By contrast, for HEX-AG, a $BF_{01} = 8.75$ indicated the opposite, with the null being almost 9 times as likely as the alternative hypothesis. Second, we compared the size of the two correlation coefficients (i.e., $r = .42$ for HEX-HH

vs. $r = .06$ for HEX-AG) using a z test for dependent correlations (Meng, Rosenthal, & Rubin, 1992). As implied by a significant difference, $z = 3.16$, $p = .002$, the correlation between trustworthiness and HEX-HH was indeed larger than its counterpart for HEX-AG. Finally, concurrently regressing the average return on both HEX-HH and HEX-AG revealed a unique impact of HEX-HH, $\beta = .43$, $p < .001$, 95% CI = [.25, .61], but not of HEX-AG, $\beta = -.05$, $p = .592$, 95% CI = [−.23, .13]. In sum, all three analyses consistently supported an influence of HEX-HH, but not of HEX-AG, which, in turn, corresponds to the unconditional kindness mechanism and is also compatible with the positive reciprocity mechanism.

However, as detailed above, an unconditional kindness mechanism also requires that the relation between HEX-HH and trustworthiness is independent of the level of trust, thus prohibiting an interaction with prior trust. By contrast, the positive reciprocity mechanism specifically necessitates said interaction. To hence test the interaction between HEX-HH and prior trust, we used a linear mixed model, regressing trustworthiness on HEX-HH (between-level predictor), trust (within-level predictor), and their interaction.[5] In line with the unconditional kindness mechanism, the model revealed a significant main effect of HEX-HH, $B = 13.47$, $p < .001$, 95% CI = [7.92, 19.01], but no interaction between HEX-HH and trust, $B = -0.02$, $p = .759$, 95% CI = [−0.16, 0.12]. To test this null interaction more conclusively, we approximated the BIC values for the two regression models (i.e., the main-effects model without the interaction term and the interaction model including the interaction term) based on their respective log-likelihood (Wagenmakers, 2007; Equation 9) and calculated the BF_{01} for the difference between the two BIC values (Wagenmakers, 2007; Equation 10). As indicated by $BF_{01} = 9.91$ (based on $\Delta BIC_{10} = 4.59$), the probability of the main-effects model given the data was almost 10 times greater than the probability of the interaction model. Overall, this pattern implies that positive reciprocity cannot account for trustworthy behavior. As displayed in Figure 1, the (positive) relation between HEX-HH and trustworthiness was indeed equivalent across trust levels.

To finally test the negative reciprocity mechanism, the same linear mixed model was used, including HEX-AG, trust, and their interaction as predictors. Contradicting the negative reciprocity mechanism, the model did not reveal an interaction between HEX-AG and trust, $B = 0.06$, $p = .481$, 95% CI = [−0.10, 0.21] (and—mirroring the results for the average return—no main effect of HEX-AG, $B = 1.96$, $p = .558$, 95% CI = [−4.64, 8.56]). Correspondingly, comparing the HEX-AG main-effects model with the interaction model revealed $BF_{01} = 8.10$ ($\Delta BIC_{10} = 4.18$), thus indicating that the probability for the main-effects model was about 8 times greater than the corresponding probability for the interaction model. Hence, the data contradicted negative reciprocity as underlying determinant of trustworthiness.

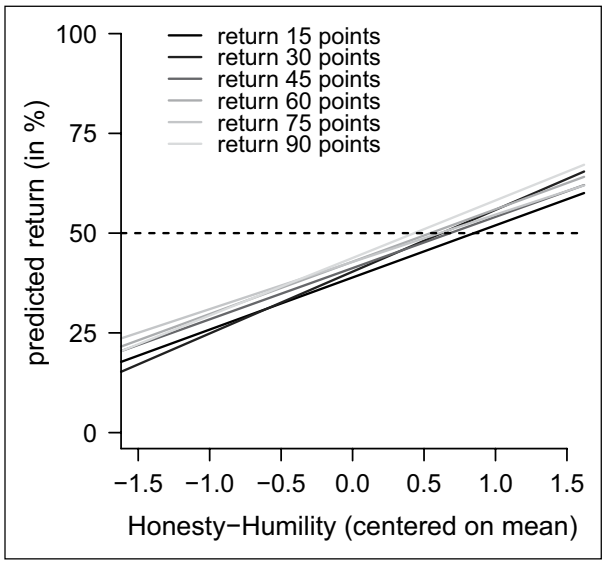

Figure 1. Predicted trustee returns for all levels of prior trust (between 15 and 90 points) depending on individual Honesty-Humility scores (centered on mean) in Study 1.

In summary, our analyses yielded a positive and consistent influence of HEX-HH on trustee returns across different trust levels. This suggests that trustworthiness is driven by dispositional tendencies of unconditional kindness. By contrast, there was no evidence favoring the positive or negative reciprocity mechanisms as we found no interactions of HEX-HH or HEX-AG, respectively, with prior trust.

However, a potential limitation of our study was that participants distributed points (rather than money) in the trust game and that only 25% of participants were incentivized. Although meta-analytic evidence on the trust game suggests that trustee behavior should be unaffected by the actual rate of incentivization (Johnson & Mislin, 2011), it nonetheless seemed prudent to replicate the above results providing monetary behavior-contingent incentives for all participants. In addition, given that the majority of participants in Study 1 were students and that 75% were female, we aimed at critically testing the results in a more heterogeneous (non-student) sample. Therefore, participants in Study 2 were recruited by an independent professional panel provider. Furthermore, we considered it important to rule out demand effects of participants' personality self-reports on subsequent behavior in the trust game. Therefore, in Study 2 we implemented a longitudinal design, separating the personality assessment from the assessment of trustworthiness in time. Finally, as outlined above, Study 2 aimed at clarifying the mixed extant evidence on the relation between FFM-AG and trustworthiness. To this end, we additionally tested which aspects of FFM-AG actually link to (or show no relation with) trustworthiness.

Table 2. Means, Standard Deviations (in Parentheses), and Bivariate Correlations (95% Confidence Intervals in Brackets) of All Focal Variables Assessed in Study 2, With Internal Consistency Reliabilities (Cronbach's α) in the Diagonal.

			Correlations		
Measure	Scale	*M* (*SD*)	HEX-HH	HEX-AG	FFM-AG
HEX-HH	1-5	3.51 (0.59)	.73		
HEX-AG	1-5	3.13 (0.52)	.45*** [.29, .58]	.77	
FFM-AG	1-5	3.62 (0.45)	.49*** [.34, .62]	.60*** [.47, .71]	.75
Return 1.50€ (in %)	0-100	49.2 (26.4)	.28** [.11, .44]	.12 [−.07, .29]	.21* [.03, .38]
Return 3.00€ (in %)	0-100	47.9 (23.6)	.21* [.03, .38]	.07 [−.11, .25]	.15 [−.04, .32]
Return 4.50€ (in %)	0-100	45.9 (22.7)	.24** [.06, .40]	.15 [−.03, .32]	.10 [−.08, .28]
Return 6.00€ (in %)	0-100	46.3 (23.5)	.26** [.09, .42]	.15 [−.04, .32]	.16 [−.02, .33]
Return 7.50€ (in %)	0-100	48.5 (22.7)	.23* [.05, .40]	.13 [−.05, .31]	.16 [−.02, .33]
Return 9.00€ (in %)	0-100	49.8 (24.0)	.27** [.09, .43]	.14 [−.04, .31]	.18* [.00, .35]
M return (in %)	0-100	47.9 (20.8)	.29** [.11, .44]	.14 [−.04, .32]	.18* [.00, .35]

Note. HEX-HH = Honesty-Humility; HEX-AG = HEXACO-Agreeableness; FFM-AG = FFM-Agreeableness.
*$p \le .05$. **$p \le .01$. ***$p \le .001$.

Study 2

Method

Materials. As in Study 1, we used the German version of the HEXACO-60 to assess the six HEXACO dimensions. In addition, the FFM factors were measured via the German 60-item NEO Five-Factor Inventory (NEO-FFI; Borkenau & Ostendorf, 2008), including 12 items for each personality factor. In both questionnaires, participants' responses were collected on 5-point Likert-type scales, ranging from *strongly disagree* to *strongly agree*.

To measure trustworthiness, we again relied on the classical trust game, with participants acting in the role of the trustee.[6] However, in contrast to Study 1, participants now played for real money (rather than points). That is, participants (as well as trustors) received an initial endowment of 3.00€ (approximately US$3.80) and were—again corresponding to the strategy method—asked to decide how much they wanted to return to the trustor for each potential (tripled) transfer (between 1.50€ and 9.00€, in 1.50€ increments). In addition, we slightly changed the response format to allow maximum transparency, such that participants received explicit information on the outcomes for themselves and the unknown trustor for all potential returns.

Procedure. Study 2 was again conducted via the Internet. Yet, to further strengthen our data compared to Study 1, we implemented a longitudinal design, separating the personality assessment from the trust game. At both measurement occasions, participants first provided informed consent and demographic information. At Time 1, they completed the HEXACO-60 and the NEO-FFI, followed by other measures not pertinent to the current investigation. At Time 2 (about 5 months later), a random subsample of participants were re-invited to a follow-up study. In this study, participants received detailed information on the rules of the trust game and indicated their returns for each potential (tripled) trust transfer as a trustee. After completing data collection, participants were randomly matched to a trustor (assessed in a separate study) to determine individual payoffs. Incentive payment (consisting of a flat fee for participation and payoffs earned in the trust game) was handled entirely (and anonymously) by the panel provider.

Participants. Following the power analysis reported in Study 1, the subsample recruited for Time 2 comprised $N = 118$ participants (51 female). Supporting the heterogeneous composition of the sample, participants' ages covered a broad range (20-66 years), with a relatively high average age ($M = 42.0$, $SD = 12.4$). Only 5.1% of participants were students, whereas about two thirds (68.6%) were in employment. Also, there was a substantial diversity in educational levels.

Results and Discussion

Unconditional kindness versus (positive/negative) reciprocity. Table 2 reports the means, standard deviations, and correlations between all variables of interest (for information on all variables, see Table S2 in the Supplemental Material). Similar to Study 1, HEX-HH showed significant (positive) correlations with trustworthiness for all trust levels individually as well as for the average return (all rs > .20, $p < .025$). For HEX-AG, in turn, no noteworthy associations with trustworthiness were apparent (all rs ≤ .15, $p > .10$). Altogether, these zero-order correlations corroborate those observed in Study 1, although effect sizes were slightly different.

To test this pattern of results statistically, we relied on the same three-step approach as in Study 1 (based on the average return across trust levels). First, we approximated the BIC and corresponding BF_{01} based on R^2 (predicting trustworthiness with HEX-HH and HEX-AG, respectively; Raftery, 1995; Wagenmakers, 2007). For HEX-HH, $BF_{01} = 0.07$

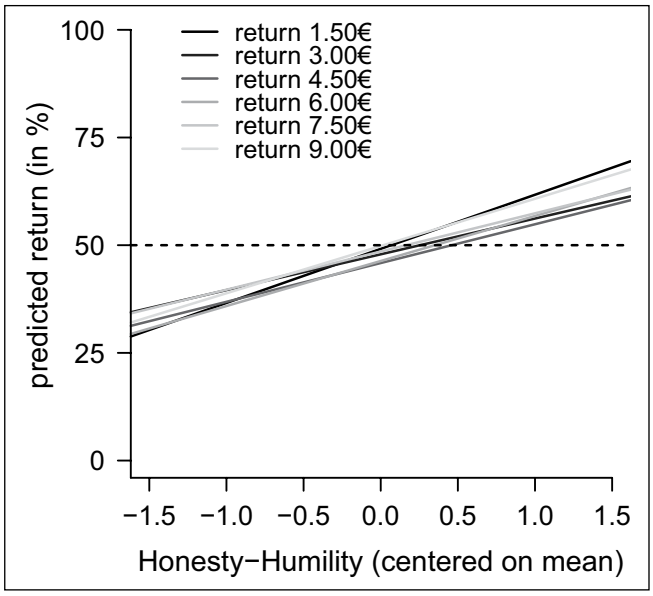

Figure 2. Predicted trustee returns for all levels of prior trust (between 1.50€ and 9.00€) depending on individual Honesty-Humility scores (centered on mean) in Study 2.

yielded substantial evidence in favor of the alternative hypothesis, being more than 14 times as likely as the null hypothesis given the data. For HEX-AG, by contrast, BF_{01} = 3.11 suggested the null hypothesis to be about 3 times as likely as the alternative hypothesis. Second, we compared the zero-order correlations between trustworthiness and HEX-HH (r = .29) and HEX-AG (r = .14), respectively (Meng et al., 1992). Unlike in Study 1, this test did not reach a conventional level of statistical significance (z = 1.49, p = .137). Nevertheless, when finally regressing the average return on both HEX-HH and HEX-AG concurrently, only HEX-HH predicted unique variance in trustworthiness, β = .28, p = .006, 95% CI = [.08, .47], whereas HEX-AG did not, β = .02, p = .830, 95% CI = [−.18, .22]. Taken together, these results largely replicate the findings of Study 1, providing evidence for the unconditional kindness and compatible with the positive reciprocity mechanism.

To further test whether the link between HEX-HH and trustworthiness is unconditional or conditional on the level of prior trust, we once more used a linear mixed model, first regressing trustworthiness on HEX-HH, trust, and their interaction (cf. Study 1). Consistent with Study 1, and in line with unconditional kindness, we found a significant main effect of HEX-HH, B = 10.03, p = .002, 95% CI = [3.92, 16.13], but no interaction between HEX-HH and trust, B = −0.28, p = .774, 95% CI = [−2.17, 1.61]. Correspondingly, comparing the main-effects model with the interaction model (cf. Study 1) yielded BF_{01} = 10.43 (based on ΔBIC_{10} = 4.69), thus indicating the main-effects model to be about 10 times as likely as the interaction model given the data. Indeed, the relation between HEX-HH and trustworthiness was again comparable for all potential trust levels (Figure 2), thus contradicting the positive reciprocity mechanism.

Finally, aiming to test the negative reciprocity mechanism, we reran the linear mixed model including HEX-AG as between-level predictor. However, the model showed no interaction between HEX-AG and prior trust in predicting trustworthiness, B = 0.60, p = .584, 95% CI = [−1.55, 2.75] (and, corroborating the results for the average return, no main effect of HEX-AG, B = 5.77, p = .114, 95% CI = [−1.40, 12.93]). Correspondingly, the BF_{01} = 9.35 (ΔBIC_{10} = 4.47) for the model comparison revealed that the main-effects model was more than 9 times as likely as the interaction model. Overall, Study 2 hence replicated the findings observed in Study 1, corroborating that trustworthy behavior can be considered an expression of unconditional kindness rather than of positive/negative reciprocity.

FFM-AG and trustworthiness. Regarding the relation between FFM-AG and trustee behavior, our results mirror previous findings in that FFM-AG showed small to medium-sized (Cohen, 1988) correlations with trustworthiness, with significant effects for some trust levels, but not for others (and r = .18, p = .045, for the average return; cf. Table 2). As reasoned above, a potential explanation for this relatively weak effect might be that only very specific aspects of FFM-AG—namely those driving unconditional kindness, as implied by the above results—actually account for trustworthiness. To test this assumption, we first regressed trustworthiness (average return) on both FFM-AG and HEX-HH, given that HEX-HH captures unconditional kindness. Indeed, only HEX-HH predicted unique variance in trustworthiness, β = .26, p = .013, 95% CI = [.05, .46], whereas FFM-AG did not, β = .06, p = .579, 95% CI = [−.15, .26]. This suggests that those aspects linking FFM-AG to trustworthiness are the ones covered by HEX-HH.[7] To test this interpretation more thoroughly, we used a mediation approach (Preacher & Kelley, 2011) in Mplus (Muthén & Muthén, 2012). Indirect effects ($a \times b$) refer to the standardized solution with the corresponding bootstrap confidence interval based on B = 1,000 bootstrap samples. As the analyses revealed, HEX-HH indeed mediated the link between FFM-AG and trustworthiness, $a \times b$ = 0.128, p = .043, 95% CI = [0.004, 0.252], but not vice versa (for FFM-AG as mediator), $a \times b$ = 0.028, p = .592, 95% CI = [−0.075, 0.132]. Overall, the results hence suggest that the relation between FFM-AG and trustworthiness can be attributed to aspects of active cooperativeness (including unconditional kindness) as covered by HEX-HH. However, as these aspects only constitute one component of FFM-AG (among several others), FFM-AG seems somewhat too broad to predict a specific behavior like trustworthiness in a satisfactory manner.

Taken together, Study 2 successfully replicated the results of Study 1 in a heterogeneous (non-student) sample with monetary behavior-contingent incentives for all participants,

using a longitudinal design. Thus, the findings support the conclusion that trustworthiness is mainly driven by unconditional kindness. Moreover, Study 2 provides clarification concerning the mixed evidence linking FFM-AG to trustworthiness: Aspects mirroring unconditional kindness are only a relatively minor component of the broad FFM-AG factor, but are well captured by HEX-HH.

Still, a limitation of Studies 1 and 2 refers to our exclusive reliance on the strategy method to assess trustworthiness. Although the strategy method has the inherent advantage of providing as much data as possible for each individual (cf. Brandts & Charness, 2011), responses are partially hypothetical in nature and do not necessarily mirror actual reactions toward another's trust. Consequently, the influence of reciprocity might be suppressed to some extent. To hence rule out that the strategy method undermined the positive/negative reciprocity mechanisms, participants in Study 3 indicated their return for one specific trust level only. Moreover, so as to provide more direct evidence on the proposed mechanisms underlying trustworthiness, we tested whether participants' expectations and evaluations regarding another's trust account for trustworthiness and further explicitly assessed participants justifications for their return decision.

Study 3

Method

Materials. Similar to Study 2, we assessed individuals' trait levels on the six HEXACO dimensions (using the German version of the HEXACO-60) and the FFM traits (using the 60-item NEO-FFI). Responses were given on 5-point Likert-type scales, ranging from *strongly disagree* to *strongly agree*.

Trustworthiness was again assessed via the classical trust game with participants acting as the trustees. However, unlike Studies 1 and 2, we now relied on the direct-response method. That is, participants only indicated their return for one specific trust level a trustor could transfer from her 6.00€ (approximately US$6.80) endowment. To ensure that all potential trust levels were almost equally covered in our data, we implemented a hypothetical design without "real" trustors and money involved. However, note that evidence supports the equivalence of trustee behavior across hypothetical and real scenarios (Holm & Nystedt, 2008). As an advantage, this procedure allowed us to systematically manipulate the trust level between participants (from 1.00€ to 6.00€, in 1.00€ increments). Correspondingly, participants were asked to imagine having received a specific transfer (tripled trust level) by an unknown other and to indicate how much they want to return.

Besides trustworthiness, we created two ad hoc scales to measure participants' evaluation of the specific trust level as well as their justification for their return decision (see Supplemental Material for individual items). The "Evaluation scale" consisted of 10 items (adjectives) in total, comprising 4 positive attributes (e.g., kind), 4 negative attributes (e.g., uncooperative), and 2 rationality-related attributes (e.g., understandable). The "Justification scale" comprised 6 items in total, with 2 items referring to each proposed mechanism (i.e., unconditional kindness, positive reciprocity, negative reciprocity). In both questionnaires, responses were given on 5-point Likert-type scales, ranging from *strongly disagree* to *strongly agree*. Analyses were based on the means of the three subscales for each questionnaire.

Procedure. Similar to Study 2, we again implemented a longitudinal design in Study 3, separating the personality assessment from the trust game. Specifically, another random subsample of participants taking part in the "pre-study" (Time 1) for Study 2 (in which participants completed the HEXACO-60 and the NEO-FFI) was re-invited to an online follow-up study (about 11 months later) by the same panel provider (excluding participants from Study 2). After providing informed consent for this follow-up study and demographic information, participants received detailed information on the rules of the trust game (as trustee). Next, they indicated the level of trust they would expect from an unknown other (between 1.00€ and 6.00€). Thereafter, participants received information on the actual (hypothetical) trust level, evaluated this trust level (using the 10-adjective Evaluation scale), and indicated how much of the tripled amount they wanted to return. Finally, participants provided reasons for their return decision (using the 6-item Justification scale). A flat fee for participation was paid out anonymously by the panel provider.

Participants. To further strengthen our conclusions, we recruited a slightly larger sample compared with Studies 1 and 2. Thus, the subsample recruited for the second measurement occasion of Study 3 comprised $N = 177$ participants (81 female). Overall, the composition of the sample was comparable to Study 2, covering a broad diversity in age (19 to 66 years, $M = 41.5$, $SD = 12.7$) and educational level. Only 7.9% of participants were students; 65.0% were in employment. Participants were almost equally distributed across trust levels (ranging between $n = 27$ and $n = 31$)

Results and Discussion

Unconditional kindness versus (positive/negative) reciprocity. Table 3 reports the means, standard deviations, and correlations between all focal variables (for information on all variables, see Table S3 in the Supplemental Material). As before, HEX-HH showed a positive (albeit weaker) relation to trustworthiness ($r = .17$, $p = .025$), which now referred to trustees' return (percentage of tripled transfer) in response to one specific trust level. For HEX-AG, a corresponding link was again absent ($r = .04$, $p = .611$). The zero-order correlations hence largely corroborated the results of Studies 1 and 2.

Table 3. Means, Standard Deviations (in Parentheses), and Bivariate Correlations (95% Confidence Intervals in Brackets) of All Focal Variables Assessed in Study 3, With Internal Consistency Reliabilities (Cronbach's α) in the Diagonal.

			Correlations			
Measure	Scale	*M* (*SD*)	HEX-HH	HEX-AG	FFM-AG	Return (in %)
HEX-HH	1-5	3.54 (0.74)	.84			
HEX-AG	1-5	3.13 (0.46)	.32*** [.18, .44]	.66		
FFM-AG	1-5	3.64 (0.45)	.56*** [.45, .66]	.53*** [.41, .63]	.73	
Return (in %)	0-100	50.5 (24.7)	.17* [.02, .31]	.04 [−.11, .18]	.05 [−.10, .20]	—
Expected trust	0-6	2.83 (1.60)	.10 [−.05, .24]	.03 [−.12, .18]	−.04 [−.18, .11]	.14 [.00, .28]
Trust evaluation: Positive	1-5	3.55 (0.91)	.08 [−.07, .22]	.19* [.04, .33]	.15* [.00, .29]	.04 [−.10, .19]
Trust evaluation: Negative	1-5	2.04 (0.96)	−.14 [−.29, .00]	−.18* [−.31, −.03]	−.22** [−.36, −.08]	−.09 [−.24, .05]
Trust evaluation: Rational	1-5	3.27 (0.88)	−.07 [−.22, .08]	.09 [−.05, .24]	.08 [−.06, .23]	−.12 [−.26, .03]
Justification: Unconditional kindness	1-5	3.53 (0.84)	.28*** [.13, .41]	.21** [.07, .35]	.17* [.03, .31]	.33*** [.19, .45]
Justification: Positive reciprocity	1-5	3.71 (0.89)	.08 [−.07, .22]	.14 [−.01, .28]	.18* [.03, .32]	.19* [.04, .33]
Justification: Negative reciprocity	1-5	1.75 (0.97)	−.18* [−.31, −.03]	−.13 [−.27, .02]	−.21** [−.35, −.06]	−.19* [−.33, −.04]

Note. HEX-HH = Honesty-Humility; HEX-AG = HEXACO-Agreeableness; FFM-AG = FFM-Agreeableness.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

To test this pattern of results statistically, we used the same three-step analytical approach as described above (now based on the return for a specific trust level). First, the approximation of the BIC and corresponding BF_{01} based on R^2 (Raftery, 1995; Wagenmakers, 2007) revealed $BF_{01} = 1.03$ for the link between HEX-HH and trustworthiness, and thus—unlike Studies 1 and 2—only inconclusive evidence (neither for nor against the alternative hypothesis). For HEX-AG, by contrast, $BF_{01} = 11.67$ indicated strong evidence in favor of the null, being almost 12 times as likely as the alternative hypothesis given the data. Second, comparing the zero-order correlations (Meng et al., 1992) between trustworthiness and HEX-HH ($r = .17$) and HEX-AG ($r = .04$), respectively, failed to reveal a significant difference, $z = 1.48$, $p = .139$. Nevertheless, in a multiple regression including both HEX-HH and HEX-AG as predictors of trustworthiness, HEX-HH predicted unique variance, $\beta = .17$, $p = .029$, 95% CI = [.02, .33], whereas HEX-AG did not, $\beta = -.02$, $p = .832$, 95% CI = [−.17, .14]. In sum, these results replicated the findings of Studies 1 and 2 only partially. Nonetheless, they still align with the unconditional kindness mechanism and are also compatible with the positive reciprocity mechanism.

To further test whether unconditional kindness or positive reciprocity accounts for the (albeit weak) link between HEX-HH and trustworthiness, we regressed trustworthiness (one return per participant) on HEX-HH, trust (between-participants), and their interaction in a multiple regression analysis. In line with unconditional kindness, we found a (one-tailed significant) main effect of HEX-HH, $\beta = .15$, $p = .053$, 95% CI = [.00, .30], but no interaction between HEX-HH and prior trust, $\beta = .00$, $p = .916$, 95% CI = [−.16, .14]. Correspondingly, comparing the main-effects-regression model with the interaction model revealed $BF_{01} = 13.23$ (based on $\Delta BIC_{10} = 5.16$), thus rendering the former 13 times as likely as the latter. Altogether, evidence once more corresponded better to the unconditional kindness mechanism than to the positive reciprocity mechanism.

Regarding the negative reciprocity mechanism, we reran the multiple regression analysis from above, now including HEX-AG as trait-based predictor. However, corroborating the results obtained with the strategy method (Studies 1 and 2), the interaction between HEX-AG and trust once more failed to explain significant variance in trustworthiness, $\beta = .02$, $p = .689$, 95% CI = [−.12, .18] (as did the main effect of HEX-AG, $\beta = .03$, $p = .672$, 95% CI = [−.12, .18]). In turn, comparing the main-effects model with the interaction model yielded $BF_{01} = 12.25$ ($\Delta BIC_{10} = 5.01$), thus implying a 12 times higher probability for the main-effects model given the data. Summing up, the trait-based evidence hence rendered the unconditional kindness mechanism most likely, thus essentially corroborating the conclusions from Studies 1 and 2.

Finally, analyses of the additional variables (i.e., expected trust, evaluations of trust, and justifications of return) provided further evidence for the unconditional kindness mechanism. First, expectations toward another's trust should reasonably influence one's interpretation of another's cooperativeness (cf. Gallucci & Perugini, 2000; see also Note 1)—as also observable in our data (Table S3 in the Supplemental Material)—and, according to a reciprocity account, affect the willingness to behave trustworthily. However, the relation between trustworthiness and expected trust level was only small (Cohen, 1988) and failed to reach statistical significance ($r = .14$, $p = .058$; Table 3). The corresponding $BF_{01} = 2.13$ implied the null hypothesis to be twice as likely as the alternative hypothesis given the data. Even smaller effect sizes emerged for individuals' evaluations of the specific trust level. That is, trustee returns were not contingent on whether participants evaluated the trust

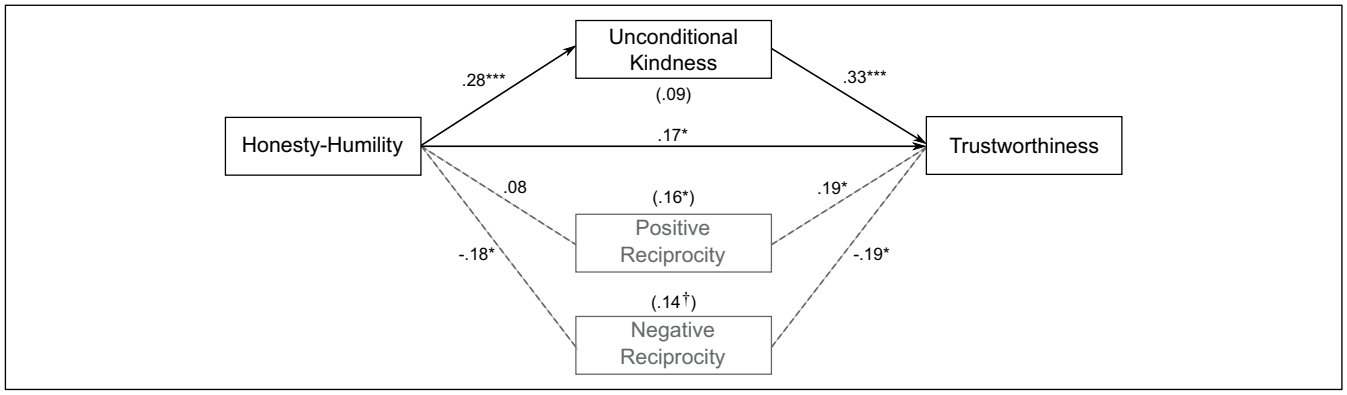

Figure 3. Mediation model of Honesty-Humility to trustworthiness with corresponding correlation and partial correlation coefficients (the latter controlling for the respective justification as mediator).
Note. Solid connections depict significant indirect effects; dashed relations depict a lack of significant indirect effects.
†$p \leq .10$. *$p \leq .05$. ***$p \leq .001$.

level positively ($r = .04$, $p = .561$, $BF_{01} = 11.21$), negatively ($r = -.09$, $p = .215$, $BF_{01} = 11.21$), or as being rational ($r = -.12$, $p = .116$, $BF_{01} = 3.81$). Moreover, participants' justification of their return decision showed the strongest correlation between trustworthiness and an unconditional kindness justification ($r = .32$, $p < .001$, $BF_{01} = 0.0006$), compared with the positive and negative reciprocity justification ($r = \pm.19$, $p = .012$, $BF_{01} = 0.54$). Correspondingly, the unconditional kindness justification was the sole significant predictor in a multiple regression including all three justification scales, $\beta = .30$, $p < .001$, 95% CI = [.15, .45]. Also, as depicted in Figure 3, the unconditional kindness justification mediated the link between HEX-HH and trustworthiness, $a \times b = 0.084$, $p = .008$, 95% CI = [0.022, 0.147], whereas both the positive ($a \times b = 0.014$, $p = .548$, 95% CI = [−0.031, 0.058]) and negative ($a \times b = 0.029$, $p = .156$, 95% CI = [−0.011, 0.070]) reciprocity justifications did not.[8] Overall, analyses of our complementary variables fit in well with the trait-based evidence from above, further supporting that trustworthiness is an expression of unconditional kindness.

FFM-AG and trustworthiness. Unlike Study 2—but converging with the mixed extant evidence—the relation between FFM-AG and trustee returns did not reach statistical significance in Study 3 ($r = .05$, $p = .505$). Correspondingly, in a multiple regression analysis concurrently considering HEX-HH and FFM-AG as predictors of trustworthiness, only HEX-HH predicted unique variance, $\beta = .21$, $p = .024$, 95% CI = [.03, .38], whereas FFM-AG did not, $\beta = -.07$, $p = .467$, 95% CI = [−.24, .11]. As such, results once more suggest that FFM-AG is somewhat too broad to account for sufficient variance in trustworthiness as a specific type of prosocial behavior.

General Discussion

The vital importance of trustworthiness for well-functioning social interactions is necessarily implied by the corresponding significance of trust. Surprisingly, however, the determinants of trustworthiness—especially in terms of (basic) personality traits—are insufficiently understood. Most prominently, two mechanisms to explain trustworthy behavior have been proposed: reciprocity (including positive and negative reciprocity) and unconditional kindness. However, purely game-theoretical evidence is mixed, and the empirical picture has remained inconclusive. The same holds for the few studies referring to basic personality traits (as conceptualized within the FFM). Most consistently, they suggest a positive, but weak and unreliable relation between trustworthiness and Agreeableness (FFM-AG). In any case, this link is insufficient to clarify the determinants underlying trustworthiness, given the broad nature of FFM-AG covering different aspects of cooperative tendencies (including positive/negative reciprocity and unconditional kindness).

To provide an enhanced understanding of the dispositional determinants of trustworthiness, we investigated the relation between trustworthiness and the HEXACO personality factors. In particular, HEX-HH and HEX-AG have consistently been linked to distinct aspects of cooperative tendencies, namely active cooperativeness (including unconditional kindness and positive reciprocity) versus reactive cooperativeness (including negative reciprocity). In turn, specific predictions on the to-be-expected relations between these two trait dimensions and trustworthy behavior can be derived for the proposed mechanisms to trustworthiness—thus disentangling unconditional kindness, positive reciprocity, and negative reciprocity through distinct hypotheses.

In three online studies, we assessed participants' trustworthiness (trustee behavior) in incentivized and hypothetical versions of the classical trust game, using either the strategy method (i.e., asking participants to indicate their trustworthiness for all potential trust levels; Studies 1 and 2) or the direct-response method (i.e., asking participants to respond to one specific trust level; Study 3). As implied by a mechanism of unconditional kindness, all studies revealed a positive link

between trustworthiness and HEX-HH, irrespective of the level of trust. Across studies, this resulted in a medium-sized average effect of $r = .28$ (sample-size weighted average correlation; Field, 2001). That is, HEX-HH showed a main effect on trustworthiness, but no interaction with prior trust—thus contradicting the positive reciprocity mechanism which inherently predicts such an interaction. HEX-AG, in turn, showed no relation to trustworthiness whatsoever (meta-analytic $r = .07$) and no interaction with prior trust, thus further contradicting that trustworthiness is determined by negative reciprocity. This interpretation was further supported by the finding that participants' expectations regarding another's trust as well as their evaluations of trust did not account for trustworthiness. In turn, participants' justification for their return decision corroborated the idea that trustworthiness is an expression of unconditional kindness. Altogether, our results are hence in line with the unconditional kindness mechanism, but cannot be reconciled with the positive or negative reciprocity mechanisms. Nevertheless, it might be worthwhile for future research to uncover potential situation-specific moderators that might render trustworthiness a conditional (reciprocal) behavior.

Overall, our findings support previous research implying that unconditional kindness is a prime determinant of trustworthiness—primarily based on a positive relation between trustworthiness and dictator game altruism (Ashraf et al., 2006; Cox, 2004; Gambetta & Przepiorka, 2014) and charitable giving (Fehrler & Przepiorka, 2013). Specifically, our studies extend the extant literature by using a trait-based approach, thus overcoming some of the inherent limitations associated with purely game-based approaches (e.g., common-method variance, desire for consistent responding). Moreover, the results—especially those clashing with the negative reciprocity mechanism—further support that trust does not correspond to a social norm which other people expect to be upheld (Bicchieri, Xiao, & Muldoon, 2011; Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014).

In addition, the results of Studies 2 and 3 replicated previous research in that they only revealed a weak link between FFM-AG and trustworthiness (meta-analytic $r = .10$). However, our results offer a reasonable explanation for this pattern: Unlike in HEX-HH, trait aspects predicting unconditional kindness are only marginally represented in FFM-AG. Correspondingly, HEX-HH mediated the link between FFM-AG and trustworthiness, but not vice versa. Similarly, HEX-HH predicted unique variance in trustworthiness beyond FFM-AG. Overall, the weak link between FFM-AG and trustworthiness thus seems to result from the misfit between the rather specific nature of trustworthiness (mainly incorporating unconditional kindness) and the broad nature of FFM-AG.

Regarding the HEXACO model in particular, the findings provide first evidence on the relation between the HEXACO dimensions and (incentivized) trust game behavior—thereby filling a gap identified in a recent meta-analytic review (Zhao & Smillie, 2015). In other words, they extend previous research linking HEX-HH and non-exploitation in the dictator game (e.g., Baumert et al., 2013; Hilbig et al., 2015) to situations in which the to-be-divided endowment is provided by another person rather than the investigator. Thereby, the results essentially corroborate the notion that HEX-HH should, by definition, drive trustworthiness (Thielmann & Hilbig, 2014).

Nonetheless, some limitations of the present studies should be acknowledged. First, the degree of interpersonal contact is obviously minimized in web-based studies. Thus, our web-based procedure might have diminished the feeling that one's own behavior is consequential for another's outcome. Although evidence suggests a high comparability of trust game behavior across web-based and lab-based studies (Holm & Nystedt, 2008), future studies might consider replicating our findings in lab-based settings. Second, due to our focus on basic and broad personality traits, we did not incorporate more specific trait scales of unconditional kindness and positive/negative reciprocity. This might be a worthwhile extension for future research. Finally, the trust game obviously represents only one specific situation eliciting trust and trustworthiness. Future studies might hence critically test the generalizability of our findings to other trust contexts.

In conclusion, the present studies provide trait-based evidence that trustworthiness is mainly an expression of unconditional kindness rather than positive or negative reciprocity. Thus, the findings contribute to the understanding of trustworthy behavior in terms of underlying personality traits and provide a valuable starting point for future research. Also, they point to the specific usefulness of HEX-HH and the higher resolution afforded by the HEXACO model to explain individual variation in trustworthiness and prosocial behavior more generally.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work reported herein was supported by grants to the second author from the Baden-Württemberg Foundation (Germany) and the German Research Foundation (HI 1600/1-1).

Notes

1. Indeed, from a game-theoretical perspective—assuming that individuals are rational utility-maximizers and thus untrustworthy—any trust transfer above zero should basically be perceived as cooperative. In turn, a trustworthy reaction necessarily implies positive reciprocity, whereas an untrustworthy reaction would simply mirror rationality (rather than implying negative reciprocity). However, whether a transfer is *actually* perceived as cooperative or uncooperative will lie in the eye of

the beholder and, for example, depend on the level of trust a trustee expects an interaction partner to place in her (Gallucci & Perugini, 2000). Correspondingly, Berg, Dickhaut, and McCabe (1995) themselves noted that trustees returning nothing "may not have *interpreted* the [trustor's] decisions as placing a trust" (p. 137, emphasis added).

2. Note that giving in the dictator game—as well as other forms of charitable giving—can also be considered in terms of fairness (cf. Forsythe, Horowitz, Savin, & Sefton, 1994) and might thus not provide an optimal measure of *pure* unconditional kindness.
3. On closer inspection, two HEXACO-Agreeableness items (items 3 and 27 of the herein used HEXACO-60; Ashton & Lee, 2009) are already conditional in nature. According to the negative reciprocity mechanism, these items should thus show an unconditional relation to trustworthiness (i.e., a simple main effect). Correspondingly, we have repeated all analyses for the two-item parcel—which basically yielded similar conclusions as will be reported below.
4. More precisely, participants were randomly assigned to the role of the trustor or the trustee. However, in what follows, we will exclusively refer to participants acting as the trustee. Specifically, trustors' responses mainly served the purpose of making the game "real" (without requiring deception).
5. The model specified the repeated trustworthiness measurements (Level 1) nested within participants (Level 2) and was estimated using the lm function of the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R. All variables were centered on the global mean. Model statistics are based on maximum-likelihood estimates. However, all models reported here and in the following were also fitted based on restricted maximum-likelihood (REML), which did not result in any noteworthy differences.
6. The trustors, to whom participants were randomly matched, were assessed in a separate study. In what follows, we will only refer to the trustees.
7. To check whether our data (based on a moderate sample size) might have over- or underestimated the differential predictive power of Five-Factor Model-Agreeableness (FFM-AG) and Honesty-Humility (HEX-HH) on trustworthiness, we compared the correlation between FFM-AG and HEX-HH (r = .49) with a larger (and representative) German sample (N = 2,027). However, supporting the validity of the current findings, the corresponding effect size in the large sample (r = .39) still fell in the 95% CI = [.34, .62] for the observed correlation.
8. Results remain similar when including the trust level as between-participants covariate in the mediation models (for details on the mediation approach, see Study 2). We refrained from calculating the same analyses for HEX-AG due to the observed null relation between HEX-AG and trustworthiness.

Supplemental Material

The online supplemental material is available at http://pspb.sagepub.com/supplemental.

References

Ackermann, K. A., Fleiß, J., & Murphy, R. O. (in press). Reciprocity as an individual difference. *Journal of Conflict Resolution*. doi: 10.1177/0022002714541854

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*, 193-208. doi: 10.1007/s10683-006-9122-4

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*, 150-166. doi: 10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*, 340-345. doi: 10.1080/00223890902935878

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*, 139-152. doi: 10.1177/1088868314523838

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4. R package* (Version 1.1-7). Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Bauernschuster, S., Falck, O., & Große, N. (2013). When trustors compete for the favour of a trustee—A laboratory experiment. *Journal of Economic Psychology*, *34*, 133-147. doi: 10.1016/j.joep.2012.09.004

Baumert, A., Schlösser, T. M., & Schmitt, M. (2013). Economic games: A performance-based assessment of fairness and altruism. *European Journal of Psychological Assessment*. doi: 10.1027/1015-5759/a000183

Becker, A., Deckers, T., Dohmen, T., Falk, A., & Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics*, *4*, 453-478. doi: 10.1146/annurev-economics-080511-110922

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*, 64-79. doi: 10.1016/j.joep.2009.10.001

Berg, J., Dickhaut, J., & McCabe, K. A. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122-142. doi: 10.1006/game.1995.1027

Bicchieri, C., Xiao, E., & Muldoon, R. (2011). Trustworthiness is a social norm, but trusting is not. *Politics, Philosophy & Economics*, *10*, 170-187. http://doi.org/10.1177/1470594X10387260

Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI). Manual (2., neu normierte und vollständig überarbeitete Auflage)* [NEO Five-Factor Inventory by Costa and McCrae (NEO-FFI). Manual (2nd ed.)]. Göttingen, Germany: Hogrefe.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*, 375-398. doi: 10.1007/s10683-011-9272-x

Chaudhuri, A., & Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, *73*, 959-985.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Hillsdale, NJ: Lawrence Erlbaum.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and NEO Five–Factor Inventory (NEO–FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences*, *12*, 887-898. doi: 10.1016/0191-8869(91)90177-D

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*, 260-281. doi: 10.1016/S0899-8256(03)00119-2

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor Model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*, 1285-1302. doi: 10.1016/j.jrp.2008.04.002

Dunning, D. A., Anderson, J. E., Schlösser, T. M., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology*, *107*, 122-141. doi: 10.1037/a0036673

Evans, A. M., & Krueger, J. I. (2009). The psychology (and economics) of trust. *Social & Personality Psychology Compass*, *3*, 1003-1017. doi: 10.1111/j.1751-9004.2009.00232.x

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*, 1585-1593. doi: 10.1016/j.jrp.2008.07.011

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160. doi: 10.3758/BRM.41.4.1149

Fehrler, S., & Przepiorka, W. (2013). Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts. *Evolution & Human Behavior*, *34*, 139-145. doi: 10.1016/j.evolhumbehav.2012.11.005

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, *6*, 161-180.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*, 347-369.

Gallucci, M., & Perugini, M. (2000). An experimental test of a game-theoretical model of reciprocity. *Journal of Behavioral Decision Making*, *13*, 367-389.

Gambetta, D., & Przepiorka, W. (2014). Natural and strategic generosity as signals of trustworthiness. *PLoS ONE*, *9*(5), e97533.

Hilbig, B. E., Glöckner, A., & Zettler, I. (2014). Personality and pro-social behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology*, *107*, 529-539. doi: 10.1037/a0036074

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S., & Zettler, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality*. doi: 10.1016/j.jrp.2014.12.004

Hilbig, B. E., Zettler, I., & Heydasch, T. (2012). Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional Honesty-Humility. *European Journal of Personality*, *26*, 245-254. doi: 10.1002/per.830

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*, 598-603. doi: 10.1016/j.paid.2012.11.008

Holm, H. J., & Nystedt, P. (2008). Trust in surveys and games—A methodological contribution on the influence of money and location. *Journal of Economic Psychology*, *29*, 522-542. doi: 10.1016/j.joep.2007.07.010

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*, 865-889. doi: 10.1016/j.joep.2011.05.007

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*, 329-358. doi: 10.1207/s15327906mbr3902_8

Lönnqvist, J.-E., Verkasalo, M., Wichardt, P. C., & Walkowitz, G. (2012). Personality disorder categories as combinations of dimensions: Translating cooperative behavior in borderline personality disorder into the five-factor framework. *Journal of Personality Disorders*, *26*, 298-304. doi:10.1521/pedi.2012.26.2.298

Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes*, *94*, 61-73. doi: 10.1016/j.obhdp.2004.03.001

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*, 267-275. doi: 10.1016/S0167-2681(03)00003-9

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology*, *49*, 710-721. doi: 10.1037/0022-3514.49.3.710

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172-175. doi: 10.1037/0033-2909.111.1.172

Moshagen, M., Hilbig, B. E., & Zettler, I. (2014). Faktorstruktur, psychometrische Eigenschaften und Messinvarianz der deutschen Version des 60-Item HEXACO Persönlichkeitsinventars [Factor structure, psychometric properties, and measurement invariance of the German-language version of the 60-item HEXACO personality inventory]. *Diagnostica*, *60*(2), 86-97.

Müller, J., & Schwieren, C. (2012). *What can the Big Five personality factors contribute to explain small-scale economic behavior?* (Tinbergen Institute Discussion Paper No. 12-028.1). Amsterdam, The Netherlands: Tinbergen Institute.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Perugini, M., & Gallucci, M. (2001). Individual differences and social norms: The distinction between reciprocators and prosocials. *European Journal of Personality*, *15*(S1), S19-S35. doi: 10.1002/per.419

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17*, 251-283.doi: 10.1002/per.474

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*(2), 93-115. doi: 10.1037/a0022658

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163. doi: 10.2307/271063

Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, *49*, 243-256. doi: 10.1026//1618-3169.49.4.243

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes [The strategy method as a tool to analyze bounded rationality in oligopoly experiments]. In H. Sauermann

(Ed.), *Beiträge zur Experimentellen Wirtschaftsforschung* [Contributions in experimental economics] (pp. 136–168). Tübingen, Germany: J. C. B. Mohr.

Thielmann, I., & Hilbig, B. E. (2014). Trust in me, trust in you: A social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *Journal of Research in Personality*, *50*(3), 61-65. doi: 10.1016/j.jrp.2014.03.006

Thielmann, I., & Hilbig, B. E. (in press). Trust: An integrative review from a person-situation perspective. *Review of General Psychology*. doi: 10.1037/gpr0000046

Thielmann, I., Hilbig, B. E., & Niedtfeld, I. (2014). Willing to give but not to forgive: Borderline personality features and cooperative behavior. *Journal of Personality Disorders*, *28*, 778-795. doi: 10.1521/pedi_2014_28_135

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779-804. doi: 10.3758/BF03194105

Yamagishi, T. (2011). *Trust: The evolutionary game of mind and society*. Tokyo, Japan: Springer. doi:10.1007/978-4-431-53936-0

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., . . . Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 20364-20368. doi: 10.1073/pnas.1212126109

Zettler, I., Hilbig, B. E., & Heydasch, T. (2013). Two sides of one coin: Honesty-Humility and situational factors mutually shape social dilemma decision making. *Journal of Research in Personality*, *47*, 286-295. doi: 10.1016/j.jrp.2013.01.012

Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review*, *19*, 277-302. doi: 10.1177/1088868314553709