Essays in Behavioral and Experimental Economics

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

vorgelegt von

Christian Koch

am 6. Mai 2014

Abteilungssprecher:	Prof. Dr. Eckhard Janeba
Referent:	Prof. Dr. Dirk Engelmann
Korreferent:	Prof. Dr. Henrik Orzen
Verteidigung:	10. Juni 2014

Für meine Eltern

Acknowledgements

First and foremost, I would like to thank my supervisor Dirk Engelmann for his guidance and support. He was a very interested and dedicated supervisor offering me lots of constructive feedback on my research. Additionally, I owe special thanks to my co-advisor Henrik Orzen and my co-author Stefan Penczyski. I benefited a lot from their comments on my work. Overall, I enjoyed the pleasant and productive atmosphere in the experimental economics group at the University of Mannheim very much.

Moreover, I thank my colleagues from the Center of Doctoral Studies in Economics (CDSE). During my entire time in Mannheim, I enjoyed the pleasant and joyful atmosphere within our graduate school. I remember lots of productive discussions with many fellow students. Although we could not always reach a consent about what constitutes "the good life" and what kind of implications the prerequisites of such a life have for economic policy, I have still perceived our discussions as very inspiring. I have been very glad to meet with Andreas Bernecker, Johannes Beutel, Robert Heimbach, Timo Hoffmann, Cornelius Müller, Alexander Paul, and Aiyong Zhu.

Additionally, I benefited from comments on my work from many other people. I will acknowledge their feedback at the beginning of each chapter separately. I also gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft.

I also want to thank those friends with whom I undertook many pleasant activities outside of academic research. I enjoyed our visits to the theater and the opera as well as our hiking trips quite a lot. Finally, I would like to thank my family and especially my parents for their continuous support.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

This dissertation is driven by the idea that a better understanding of human behavior and human nature is often a prerequisite for a better understanding of many economic problems. I hope that considering psychological insights when anaylzing economic problems increases the predictive power of the economic framework. This dissertation consists of three chapters that are written such that they can be read independently. All chapters, however, develop around two central themes in behavioral and experimental economics that can been seen as behavioral economic deviations from standard economic theory: social preferences and bounded rationality. In all chapters, the experimental method is used to analyze economic questions. Chapter 2 analyzes a particular cause for social preferences, and Chapter 3 investigates a particular implication of social preferences. Finally, Chapter 4 discusses to what extent subjects show boundedly rational behavior in an auction setting.

In Chapter 2, I analyze a specific reason why people might behave pro-socially. Already Aristotle has claimed that there is a special link between *long-run* well-being and pro-social behavior. I design an experiment to investigate this link and find evidence that there seems to be a crucial connection between long-run well-being and pro-social behavior. In Chapter 3, I investigate which implications social preferences might have for the labor market. I analyze to what extent social preferences - or more precisely the reference dependence of social preferences - provide an explanation for wage rigidity, and my experimental data indeed supports this idea. Chapter 4 is joint work with Stefan Penczynski. We analyze whether people have difficulties with contingent reasoning on hypothetical events and whether this problem is at the origin of the so-called winner's curse. Our experimental data underpins this conjecture.

The appendices for each chapter that include supplementary material such as additional graphs and additional analysis as well as experimental instructions can be found at the end of this dissertation. References from all three chapters are collected in one bibliography.

1.1 Chapter 2: The Virtue Ethics Hypothesis: Is there a Nexus between Virtues and Well-Being?

This chapter starts with the observation that many people seem to be motivated not only by material self-interest but also by the material payoffs of others. In laboratory experiments, many participants for example share their initial endowment with others, trust and cooperate with each other or punish free-riders (e.g. dictator game, trust game and public goods game). Additionally, it has even been claimed that a country's economic prosperity depends on the willingness of its citizens to behave pro-socially: social capital theory has argued that different levels of trust in societies explain differences in economic prosperity across countries (see e.g. Fukuyama 1995; Knack and Keefer 1997; Algan and Cahuc 2010).

Moreover, a common finding of the experimental literature is that we observe heterogeneity in pro-social behavior (see e.g. Fischbacher et al. 2001; Bohnet et al. 2006): Some people behave altruistically, others do not, leading to the question why some people behave pro-socially (whereas other people do not). In Chapter 2, I analyze one particular explanation why (at least) some people behave pro-socially. Aristotle (and others) have suggested that there is a nexus between virtues and long-run (enduring) well-being and that only virtuous or pro-social behavior, not material affluence, makes people happy in the long-run. I try to investigate whether there is indeed a decisive connection between virtuous or pro-social behavior and well-being, adapting an experimental design proposed by Konow and Earley (2008).

In my design, subjects first answer an elaborated well-being questionnaire that asks questions about long-run and short-run well-being and which makes use of two different well-being approaches: hedonic vs. eudaimonic well-being. In a very broad sense, the hedonic approach defines well-being as the balance of pleasure vs. pain whereas the eudaimonic approach defines well-being in terms of whether subjects have a purpose in life. Afterwards, subjects play a set of classical economic games (dictator game, sequential prisoner's dilemma, mini-ultimatum game etc.) in which they can show either pro-social or egoistic behavior as well as antisocial behavior.

Overall, I find favorable correlations between well-being measures and pro-social behavior in the games and analyze various possible hypotheses regarding the underlying causality. In my setting, pro-social behavior is to a higher extent correlated with long-run well-being than with short-run well-being. More precisely, my data is mostly in line with the hypothesis that virtuous or pro-social behavior is both a long-run cause as well as a short-run effect of (long-run) *eudaimonic* well-being. Additionally, pro-social behavior in different settings is highly correlated with each other and subjects who behave pro-socially more frequently also report higher well-being. Hence, well-being and virtuous or pro-social behavior are connected with each other to the extent that we think about well-being not

just in terms of pleasure and pain (as the hedonic well-being approach does) but in terms of whether people have something like a purpose in life or whether they are striving for self-fulfillment. To this extent, I find evidence in favor of a nexus between virtues and well-being that may also help us to explain why at least some people behave pro-socially.

1.2 Chapter 3: Do Reference Points Lead to Wage Rigidity? Experimental Evidence

This chapter starts with the observation that wages seem to be downward rigid during recessions (Fehr and Goette 2005; Bewley 1999). In the literature, different explanations for this wage rigidity phenomenon have been suggested (e.g. implicit contract theory, insider-outsider theory, efficiency wage theory etc.), but questionnaire studies (e.g. Blinder and Choi 1990; Campbell and Kamlani 1997; Bewley 1999) interviewing manager about the phenomenon mainly suggest that fairness considerations play a key role: Firms seem to be constrained in their wage cuts by workers' fairness considerations that make them fear adverse effects on work morale. But why do workers perceive wage cuts as unfair even though firms have a good reason to cut wages, namely their profit decrease during a recession?

One reason could be that workers' fairness considerations are reference-point dependent and that contracts concluded before recessions could serve as such a reference point. Or in other words, what workers perceive as a fair wage in recession may just depend on the wage they previously have received out of recession. This reference dependence of fairness considerations might imply that workers perceive wage cuts as reference point violations. Workers might punish these violations by lowering their effort. Hence, firms could anticipate this behavior, leading to more rigid wages than without reference-dependent fairness considerations.

My analysis provides the first rigorous experimental test whether the dependence of fairness considerations on reference points indeed provides one explanation for wage rigidity. I am using an experimental labor market (gift-exchange game) in which firms face recessions in some periods. I mainly compare two different treatments: In one treatment, (initially concluded) contracts may serve as a reference point whereas this is not the case in the other treatment. In the first treatment, firms and workers initially conclude a contract before a recession occurs and this initial contract may then serve as a reference point if wages have to be renegotiated in case a recession actually occurs. In the second treatment, such kind of initial contracts do not exist and contracts are only concluded after it has been determined whether a recession has occurred or not. Hence, in this treatment contracts cannot serve as reference points.

My main finding is that wage cuts have only half the size if (initially concluded)

contracts potentially serve as reference points compared to the situation in which this is not the case. So, workers perceive the recession wage differently when a previous out-of-recession wage can serve as a reference point compared to when this is not the case. Hence, the reference dependence of fairness considerations leads to wage rigidity in the laboratory, which suggests that it is indeed one possible explanation for the field phenomenon.

1.3 Chapter 4: The Winner's Curse - Contingent Reasoning & Belief Formation

This chapter starts with observation that it has recently been argued that people have problems with contingent reasoning on hypothetical events. Based on experiments with computerized opponents, Charness and Levin (2009) have claimed that this cognitive difficulty might be at the origin of the winner's curse - the phenomenon that people in common value auctions systematically overbid, which typically results in severe losses. In this setting, subjects often fail to condition their behavior on the critical future event of winning the auction.

The authors, however, have come to this conclusion only indirectly by observing that subjects still tend to overbid in an individual choice problem similar to the winner's curse (acquiring-a-company game with computerized opponents). Hence, subjects fall prey to the winner's curse although they do not face the strategic uncertainty accompanied with human opponents that is typically part of a standard auction setting. Because it is not clear whether people's cognitive strategies differ between settings with human and computerized opponents, it is a priori also unclear whether the observations of Charness and Levin (2009) really provide sufficient evidence to conclude that problems with contingent reasoning lead to the winner's curse in standard common-value auctions.

We investigate this question by comparing decision makers' behavior in a simplified common-value *auction game* with their behavior in a *transformed game* that does not require contingent reasoning on the future event of winning the game. Simplifying the standard common-value auction setting by reducing the number of players and restricting the values of the information signals that subjects receive in this setting allows us to construct a number choosing game (framed as an auction) that is strategically very similar to the auction setting. The rules of this transformed game follow directly from the simplified auction game and already contain the contingent reasoning on the future event of winning the game, which subjects in the original auction setting have to do on their own. In both settings, the simplified auction and the transformed setting, subjects first play against human opponents, but we also implement a setting in which subjects face a computerized opponent with a known strategy, similar to the setup of Charness and Levin

(2009).

We have two main findings. First, in the transformed game in which the contingent reasoning problem was switched off, subjects are to a larger extent able to avoid the winner's curse compared to the original auction game. Importantly, this result holds not only in an environment with computerized opponents but also with human opponents. Second, but even when the contingent reasoning problem is switched off, many subjects fail to avoid the winner's curse when facing human opponents compared to computerized opponents. Hence, the problem of belief formation seems to be another obstacle in avoiding the winner's curse. Overall, we conclude that Charness and Levin (2009) seem to be right that the cognitive difficulties associated with contingent reasoning on hypothetical events are a serious problem in common-value auctions. Nonetheless, the problem of belief formation in this setting can also not be underestimated.

Chapter 2

The Virtue Ethics Hypothesis: Is there a Nexus between Virtues and Well-Being?[1]

2.1 Introduction

In economics, social preferences are receiving increased attention, to a large part driven by results of laboratory and field experiments. Many people seem to be motivated not only by material self-interest but also by the material payoffs of others (see e.g. Fehr and Fischbacher 2002). In laboratory experiments, many participants for example share their initial endowment with others, trust, and cooperate with each other or punish free-riders (e.g. dictator game, trust game, and public goods game). This kind of pro-social behavior has been used to analyze among other things charitable giving, the provision of public goods, and the emergence of efficiency wages (see e.g. Karlan and List 2007; Ledyard 1995; Fehr et al. 1993). Although there is some dispute about to what extent experimental results about social preferences can be generalized to the field (see e.g. Levitt and List 2007; Camerer 2011), it is claimed that a country's economic prosperity depends on the willingness of its citizens to behave pro-socially. In political science for example, social capital theory has argued that different levels of trust in societies explain differences in economic prosperity across countries (see e.g. Banfield 1958; Fukuyama 1995; Knack and Keefer 1997). More recently, economists have confirmed that there is indeed a causal

[1]For helpful comments and suggestions, I thank Andreas Bernecker, Dirk Engelmann, Robert Heimbach, James Konow, Henrik Orzen, Alexander Paul, and Stefan Penczynski. I also received helpful comments from participants at seminars in Mannheim, Heidelberg, ESA European Conference 2012 (Cologne), Mainz Workshop on Behavioral Economics 2012, Gesellschaft für experimentelle Wirtschaftsforschung 2012 (Karlsruhe), Public Happiness - HEIRS conference 2013 (Rome), EEA Annual Meeting 2013 (Gothenborg), Verein für Socialpolitik Annual Meeting 2013 (Düsseldorf). The idea of this chapter is partially based on my dissertation proposal (Koch 2011). Importantly, no experimental sessions were run in this proposal and the experimental design also has changed substantially.

effect of trust on growth (see Tabellini 2008; Algan and Cahuc 2010). The social capital literature argues that trust is a key ingredient in almost any commercial transaction and that the development of a successful market economy depends on the evolution of trust (cf. Arrow 1972, p. 357). But trust of course can only develop, if people expect others to cooperate. Cooperation however remains fragile, if free-riding behavior is not sufficiently punished.

A common and arguably not unsurprising finding of the experimental literature is that we observe heterogeneity in pro-social behavior (see e.g. Fischbacher et al. 2001; Bohnet et al. 2006): Some people trust, others do not, some people cooperate, others do not etc. This naturally leads to the following questions: Why do some people give, trust, cooperate, and punish unfair behavior? And why do other people instead behave selfishly? And, if one knows the answer, how can the fraction of pro-socially behaving citizens in a society be increased?

So far, the literature does not provide conclusive answers to these questions. In this paper, I will analyze one particular explanation why some people behave pro-socially that we may call the *Virtue Ethics Hypothesis (VEH).* This hypothesis has first been proposed by Aristotle (1987) in his *Nicomachean Ethics* and is nowadays at the center of *virtue ethics*, one branch of normative ethics besides Kantianism and consequentialism (see e.g. Anscombe 1958; Foot 1978). Aristotle claims that there is a nexus[2] between *long-run* (or *enduring*) well-being and virtues and suggests that well-being arises from a life of virtue, where virtues are seen as "acquired character-traits or dispositions that are judged to be good."[3] The Aristotelian idea clearly suggests one causal direction: Repeated acts of virtuous behavior (arising from a moral character) lead to long-run well-being. Causality, however, could of course also run the other way, namely, higher long-run well-being could also lead to more virtuous behavior (fostering the development of a virtuous character trait). The crucial point of the VEH is that it claims that long-run well-being and virtuous behavior are decisively connected, either via the first or via both lines of causality. Psychological evidence in favor of this hypothesis has recently been reviewed by Kesebir and Diener (2013).

Assuming that people strive for well-being, such a nexus would provide an explanation for virtuous behavior: People behave virtuously because it makes them happy in the long

[2]In principle, the term *nexus* means nothing more than *relationship* or *correlation.* One reason not just to use the more simple expression *relationship* is that in the philosophical literature the term *nexus* is used in connection with well-being and virtues. Additionally, this literature has claimed this nexus between virtues and well-being to be a causal one.

[3]cf. Bruni and Sugden (2013, p.143), also for more details about Virtue Ethics. An important point the authors state is that virtue ethics does not focus so much on the right actions as consequentialism and deontology potentially do but is more concerned about a person's moral character. This means that virtues are seen as good per se and not because they induce certain action. More precisely, in virtue ethics, actions are only seen as good to the extent that they "are in character for a virtuous person." Importantly however, we can not directly observe a person's moral character. Hence, we have to observe his or her actions.

run. But why do we observe heterogeneity? Why do some people not behave virtuously although it would make them happy? We may conjecture that the degree of people's ethical maturity (defined as an individual's insight into the relationship of virtuous behavior and well-being and their development of a suitable character trait) decides whether a person behaves virtuously or not. Mature people know that virtuous behavior increases their well-being whereas less mature people are unaware of this insight and do not behave virtuously. Additionally, the philosophical literature claims that one cannot make use of this nexus directly, but only by acquiring a suitable character trait. Or in other words, the right motivation matters: Behaving virtuously in order to help others will make you happy but showing virtuous behavior in order to increase your own well-being will not make you happy. Only overcoming self-centered behavior (and acquiring suitable character traits) leads to happiness.

I will analyze this Virtue Ethics Hypothesis based upon an experimental design of Konow and Earley (2008) (henceforth KE). The authors (p. 2) examine a related question, the so-called *hedonistic paradox*: This means that "the person who seeks pleasure for him- or herself will not find it, but the person who helps others will." Crucially however, KE interpret this paradox not in the sense that there is general connection between well-being and virtuous behavior (as suggested by the VEH) but they focus more narrowly on the relationship of generosity and well-being. They combine a dictator game with an extensive well-being questionnaire including several well-being measures and find a favorable relationship between long-run well-being and generosity in the dictator game. Importantly however, the non-strategic structure of the dictator game surely provides a suitable setting to measure generosity but it (deliberately) excludes any kind of reciprocal interactions that play a key role in typical economic interactions.

Hence, the first contribution of my paper is to analyze whether a nexus between well-being and virtuous behavior also exists in economically more important settings in which reciprocity plays an important role: I will basically focus on those aspects of pro-social behavior already introduced before: trust, cooperation (positive reciprocity), and punishment of unfair behavior (negative reciprocity).[4] These behavioral patterns seem to at least partially reflect aspects of one of the four cardinal virtues: Justice. Other important virtues such as e.g. the other three cardinal virtues prudence, courage, or temperance are, however, beyond the scope of this article. Whether the finding of KE extends to other games is indeed an open question: In their within-subject analysis of pro-social behavior, Blanco et al. (2011) show that their subjects' behavior in their (modified) dictator game is hardly correlated with subjects' behavior in other games. Hence, it is unclear whether KE's finding are also valid in other settings.

For my analysis, I use a within-subject design, in which subjects first answer several

[4] I will use the terms *cooperation* and *positive reciprocity* as well as *punishment* and *negative reciprocity* as synonyms.

well-being questions and then play six different cooperation games measuring different aspects of virtues (or pro-social) behavior. The well-being questionnaire has two main features. First, it covers both *long-run* as well as *short-run* well-being. Second, it covers two different well-being concepts, *eudaimonic* vs. *hedonic* well-being, which will be explained in detail in the next section. I use the following games: dictator game, sequential prisoner's dilemma, mini-ultimatum game, joy-of-destruction game, and two third-party punishment games. These games provide subjects with the opportunity to trust and to behave positively or negatively reciprocal and show spiteful/antisocial preferences.[5] Additionally, the games distinguish between different forms of punishment: Punishment by an involved party, so-called second-party (SP) punishment, and punishment by an uninvolved party, so-called third-party (TP) punishment. Hence, a second contribution of my paper is to analyze in more detail under which circumstances punishment of unfair behavior can been seen as personally beneficial in the sense that it increases or decrease the punisher's well-being. So far, the literature (see e.g. Gächter et al. 2008; Herrmann et al. 2008; Abbink et al. 2010) has mainly discussed whether punishment is socially beneficial, that is whether punishment increases overall (monetary) efficiency in the sense that monetary gains by higher cooperation levels induced by punishment exceed investments in punishment. Importantly, however, socially beneficial punishment might still not be desirable when even those people who punish (and not only those who are punished) suffer from a severe decrease in their well-being.

In line with KE, I am first of all interested in the following questions: Do virtuously behaving people report on average greater well-being in my different settings? That is, do those experimental participants who give, trust, cooperate, or punish unfair behavior report on average greater well-being than those who behave non-virtuously? If the answer is affirmative, a natural question is what kind of causal relationship underlies these findings. For this purpose, I analyze to what extent my data is in line with several competing hypotheses about what leads to virtuous behavior. For these hypotheses, the two different well-being concepts are used. Additionally, the questionnaire not only asks about long-run but also about short-run well-being. The question then is whether long-run or short-run well-being, whether hedonic or eudaimonic well-being are decisively connected with virtuous behavior. Finally, my within-subject design allows a third contribution. I am able to analyze whether people consistently act virtuously across different settings and whether those people who act virtuously more frequently also report higher well-being. The VEH states that well-being arises from a life of virtue triggered by a suitable character trait and that hence such behavior in different settings should be strongly correlated and more frequent virtuous behavior should lead to higher well-being.

Results suggest that there are positive correlations between generosity, trust, positive

[5]Importantly, not enough subjects showed spiteful/ antisocial preferences. Therefore, I cannot analyze this type of non-virtuous behavior fruitfully.

reciprocity (cooperation) and long-run well-being measures, whereas correlations between generosity, trust, positive reciprocity and short-run well-being are much more limited. In line with KE, the experimental evidence is most strongly in line with the hypothesis that virtuous behavior is both a long-run cause as well as a short-run effect of a specific type of long-run well-being, called *eudaimonic well-being*. Hence, KE's finding for generosity is also valid for the economically more important patterns of trust and cooperation. Additionally, there seems to be a crucial distinction between SP- and TP-punishment. Only the more altruistic TP-punishment is positively related with well-being. Finally, virtuous behavior in different settings is highly correlated with each other and subjects behaving virtuously more frequently also report higher well-being. Overall, Aristotle's main insight into the relationship between well-being and virtuous behavior seems to be right. There seems to be a nexus between virtues and (eudaimonic) well-being.

My paper is related to an increasing experimental literature combining laboratory experiments and happiness research (see e.g. Charness and Grosskopf 2001; Ifcher and Zarghamee 2011). In connection with virtuous or pro-social behavior, nearly all existing experimental studies besides KE have focused on the relationship of pro-social behavior and short-run well-being or mood, whereas the Virtue Ethics Hypothesis focuses on the connection with long-run well-being. Asking short-run well-being questions, Bosman and Winden (2002) look at the power-to-take game, Kirchsteiger et al. (2006) consider the gift exchange game, Becchetti and Antoni (2010) analyze the trust game, and Konow (2010) looks at the dictator game. Overall, this evidence supports the view that short-run well-being or emotions can play a role in virtuous behavior. Additionally, neuroeconomic studies support this view by analyzing people's brain activity: They find that certain brain activities cause (or are at least related to) pro-social behavior (see Fehr et al. 2005; de Quervain et al. 2004; Kosfeld et al. 2005; Knoch et al. 2006). These results are obviously interesting in themselves. They help us to understand why people behave in a certain manner. But they do not provide us with an explanation why people differ in their behavior, that is, why some people show a particular brain activity and others do not, or why some people are driven by certain emotions and others are not. Here, the VEH postulates that the insight into the relationship between virtues and well-being makes the difference for people's behavior.

Additionally, there is a literature in psychology that relates pro-social behavior in cooperation games and personality traits (see Kurzban and Houser 2001; Hirsch and Peterson 2009; Fleming and Zizzo 2011; Volk et al. 2011). However, these studies of course do not explain why some people develop certain personality traits that potentially make them behave pro-socially. In line with the VEH, it has been argued in evolutionary biology (Gintis 2000, Gintis et al. 2003, p.153) that strong reciprocity as "a predisposition to cooperate with others and to punish those who violate the norms of cooperation [...] when it is implausible to expect that these costs will be repaid" is an evolutionary stable strategy.

Importantly however, I argue in my paper that virtuous behavior has a positive impact on well-being beyond monetary gains. Finally, there are some empirical economic papers that try to investigate whether a nexus between virtues and well-being exists (see James 2011; Guven 2011). These studies found evidence in favor of such a nexus. Becchetti et al. (2013) extend this empirical analysis by also looking at a eudaimonic well-being measure. Additionally, the authors do not only look at actions but also at the motivations underlying pro-social behavior and find that pro-social behavior only leads to increased well-being when combined with other-regarding motivations. One problem of such empirical studies, however, is that they basically have to rely on survey questions about virtuous behavior. A major advantage of the experimental approach is that it provides reliable data on ethical behavior. In my complementary experimental analysis, ethical behavior is incentivized and hence comes at a cost. Additionally, in typical surveys, only one well-being concept is used (and also Becchetti et al. 2013 rely on one single eudaimonic question). In contrast, in my well-being questionnaire, two different well-being approaches are used.

The remainder of this chapter is structured as follows: Section 2.2 provides some background information regarding happiness research and the measurement of well-being. Section 2.3 outlines the experimental design. Section 2.4 presents the results, and Section 2.5 concludes.

2.2 Background: Well-Being

The economics of happiness has developed into a large and diversified field. In the following, I will first focus on the distinction between the two different well-being approaches that are used in my questionnaire: Hedonic and eudaimonic well-being.[6] Afterwards, I will provide the reader with a eudaimonic interpretation of the Virtue Ethics Hypothesis. For more general topics within this literature the reader is referred to respective survey articles and books (e.g. Frey and Stutzer 2002; Diener and Seligman 2004; Layard 2005.)

In happiness research, two main competing well-being approaches have emerged: Hedonic and eudaimonic well-being. Ryan and Deci (2001) (and also KE, Sec. 2.3) provide an excellent overview of the two different research traditions, on which my remarks are also based. Most importantly, the two approaches differ in their perception of what well-being really is about. The *hedonic school* (see e.g. Kahneman et al., eds 1999; Layard 2005) follows a distinct empirical approach and focuses on well-being in the form of *happiness* - that is, well-being is defined in terms of the antagonism of pleasure versus pain. This means that well-being is considered as an outcome variable typically measured by subjective well-being (SWB): People rate their well-being by themselves according to

[6] In the literature the terms *well-being* and *happiness* are often used as synonyms. In the following, I will try to refer to *well-being* as the more general term and to *happiness* only in connection with the hedonic well-being approach.

their own standards. The *eudaimonic school* (see e.g. Ryff 1989; Ryff and Singer 2008) is more theoretical. It examines well-being in form of *eudaimonia* (direct translation "good spirit", meaning "human flourishing") that arises from a process of human growth and self-fulfillment. It tries to theoretically establish criteria for well-being and looks to what extent people match these criteria. These criteria are derived from concepts such as human growth, self-fulfillment or human flourishing. Eudaimonic well-being can be measured by psychological well-being (PWB): The idea of this concept is that psychological scales define criteria for the fully-functioning individual "living a life rich in purpose and meaning" (Ryff and Singer, 2008, p.1). So, put in highly simple terms, the most important difference between SWB (hedonic approach) and PWB (eudaimonic approach) is that the former approach is about maximizing pleasure whereas the later approach is about something like a purpose in life.

Overall, the eudaimonic literature is fairly diverse. KE (p. 7), however, suggest two consistent features of the different contributions of the eudaimonic literature: First, SWB (as the measure of hedonic well-being) is only seen as a "favorable by-product" of PWB (as the measure of eudaimonic well-being). Or in other words: Having a purpose in life potentially also leads to pleasurable feelings. Second, this literature argues that only certain kinds of attitudes and behavior are considered to foster PWB. Aristotle would argue that long-run well-being (PWB) only arises from a life of virtue. A more modern interpretation is provided by Sheldon and Kassner (1995): The authors distinguish between intrinsic and extrinsic goals (self-acceptance, affiliation and community feeling versus financial success, popularity, and attractiveness). People following the former goals are found to report higher values of PWB and SWB.

In order to formalize these ideas, KE (Sec. 2.4) provide an economic interpretation of PWB, using a framework of Graham and Oswald (2010) who introduce the stock-flow concept into the happiness literature. In KE's interpretation of this model, people have a *stock* of psychological well-being (PWB). We may think of this stock as a set of acquired personality characteristics to which one can only add via certain behavior. This stock of PWB then produces a *flow* of psychological resources and we may think of this flow as some kind of coping resources that determine to what extent people are able to cope with their life circumstances. In the model, these psychological resources can then be used for two purposes: First, these resources can be used to generate pleasure (hedonic happiness - SWB); second, they can be used to invest in the stock of PWB via specific behavior, especially virtuous behavior.[7] In the steady-state of the model, a higher stock of PWB results in higher average returns in happiness (SWB) and a higher level of investment (virtuous behavior). Figure 2.1 illustrates this concept.

[7]Importantly, as already mentioned in the Introduction, the philosophical literature suggests that investing in the stock of PWB only works with a truly virtuous motivation. So the investment has to be undeliberate to the extent that it does not arise from self-centered thoughts.

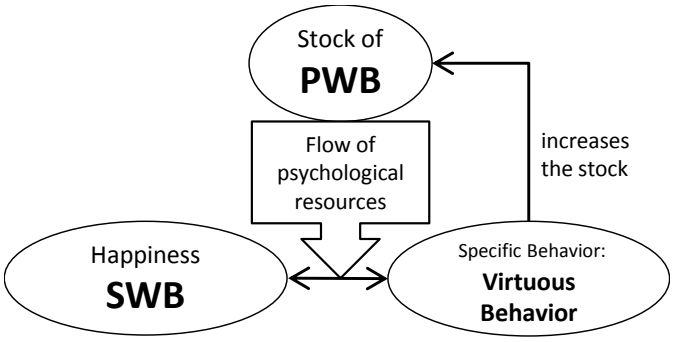


Figure 2.1: Stock-Flow Model of Well-Being

Probably the reader should not take the model too literally. Strictly speaking, it would suggest that e.g. helping others is not directly rewarding in itself because such behavior would first of all only increase the level of PWB, and hence only indirectly lead to more hedonic happiness (SWB). Anecdotal evidence might on the contrary suggest, that many people for example try to help the poor because helping others as such makes them directly happy. Nonetheless, the concept of *virtuous behavior* seems to imply that there is a trade-off between your own and other people's well-being. The model captures this trade-off in the strictest way: Either you are happy or you behave virtuously. Or in other words, either you behave self-centeredly and help yourself or you overcome your egotism and help others. And if you behave virtuously, you will only get happy in the long-run.[8]

Considering the two different approaches of well-being and their relationship towards each other, we have come to what can be called a eudaimonic interpretation of the Virtue Ethics Hypothesis (VEH). That is, a VEH in which eudaimonic well-being (or its measure PWB) plays the key role and not hedonic well-being (or its measure SWB). In this sense, the core implication of the model is the following: Virtuous behavior is a *long-run* cause of PWB and a *short-run* effect of it. This means that continuous virtuous behavior is one cause of a high level of PWB. And subsequently, a high level of PWB causes virtuous behavior (and hedonic happiness) by providing the necessary psychological resources for such behavior.

Does the literature provide any evidence in favor of this model and its implications? As already outlined by KE (p. 8), positive psychology seems to provide evidence for the claim that virtuous behavior is a long-run cause of long-run well-being (PWB): In longitudinal studies, performing acts of kindness has been shown to cause higher enduring well-being (Sheldon and Lyubomirsky 2006; Lyubomirsky et al. 2005; Lyubomirsky et al. 2011). Virtuous behavior as pro-social behavior in laboratory experiments can be interpreted as a form of kind behavior, at least if giving, trusting, and cooperating are considered. However for punishment, this is not so clear. The crucial question seems to be to what

[8]Probably it is noteworthy, that this kind of strict interpretation might be especially suitable for my laboratory setting because participants do not know to whom they behave kindly, restricting a potential warm glow induced by virtuous behavior.

extent punishing others can be considered as an act of kindness rather than an act of mere retaliation or revenge. In my design, I distinguish between SP- and TP-punishment. I will discuss different motivational forces for these forms of punishment in more detail in Section 2.3.4.

To sum it up, the eudaimonic interpretation of the VEH, introduced in this section, will be my major hypothesis concerning the relationship of virtuous behavior and well-being. As outlined in this section, this hypothesis postulates both lines of causality. First of all, I think it is important to analyze whether there is indeed a relationship between PWB and virtuous behavior, irrespective of causality. Regarding causality however, I do not perform a longitudinal study that manipulates long-run well-being as some studies in psychology have done. Hence, I will focus on the question whether virtuous behavior appears to be a short effect of a high level of PWB. That is, participants enter the laboratory with a level of PWB that is fixed in the short-run. I am then interested in whether subjects higher in PWB have a higher probability to behave virtuously. In line with the outlined stock-flow model of well-being, this line of causality (PWB causes virtuous behavior) seems plausible and hence I will phrase my hypotheses in line with this particular direction of causality. Nonetheless it is important to note that long-run well-being (in contrast to short-run well-being) cannot be experimentally varied easily and hence there is no exogenous variation in long-run well-being in my data. This means that a possible finding that virtuous behavior seems to be a short-run effect of PWB could also fully be driven by the fact that virtuous behavior is a long-run cause of PWB, as shown by psychologists. But even in this case, there is still a crucial connection between virtuous behavior and long-run well-being.

Regarding the question whether more frequent virtuous behavior should be associated with higher well-being, the stock-flow model suggests the following: In the long run, more frequent virtuous behavior is seen as a higher level of investment in the stock of PWB which will lead to a higher steady state level of PWB. In the short run, a higher stock of PWB produces more psychological resources that can then be used for more frequent virtuous behavior. Additionally, the psychological literature (Lyubomirsky et al., 2005; Lyubomirsky et al., 2011; Sheldon and Lyubomirsky, 2006) claims that it is really the frequency not the size of a kind act that matters. It seems that giving a small amount twice is better than giving a large amount once. So overall, I conclude that more frequent virtuous behavior should be related with higher well-being.

2.3 Experimental Design

In order to analyze the relationship between virtues and well-being, I implemented an experiment in which subjects first answered an extensive well-being questionnaire with more than 100 questions and then played six cooperation games. They did not receive

any feedback regarding their game decisions until the end of the experiment. Additionally, participants answered questions on their instantaneous mood after the first game, after the second game, and after the last game. At the end, they made decisions about simple lotteries (risk elicitation) and completed a concluding questionnaire.

2.3.1 The Well-Being Questionnaire

Before playing the games, subjects answer an extensive well-being questionnaire that contains all the well-being measures used by KE (plus one additional measure). The questionnaire's basic structure is the following: It asks questions about hedonic and eudaimonic well-being and distinguishes between long-run and short-run well-being. It consists of ten measures of subjective well-being (SWB - hedonic approach) and three measures of psychological well-being (PWB - eudaimonic approach). SWB has basically two components: an affective and a cognitive-evaluative component. Additionally, we can distinguish between long-run and short-run SWB. In contrast, PWB is considered as a (long-run) personality trait that is fixed in the short-run.[9]

For long-run SWB, there are seven measures: An overall happiness question (OH: "Overall, how would you describe yourself") and two similar questions about the "highest" (HH) and the "lowest" (LH) happiness level subjects have experienced. To measure the long-run affective SWB component, four measures are used: the Positive and Negative Affect Schedule (PAS and NAS) by Watson et al. (1988) and the five Positive Affect (PA) and Negative Affect (NA) items by Bradburn (1969). Short-run SWB is measured by two instruments: A single now happiness question (NH: "Right now, how would you describe yourself"), which measures the cognitive-evaluative component, and the Mood Index (MI) by Batson et al. (1988), which measures short-run affect or mood. NH and MI are measured at the very beginning, after the first, after the second, and after the last game. By subtracting the score of NH/MI at one point in time from the score of NH/MI at an earlier point in time, I will examine the change in current happiness (NHD) and the change in mood (MID) followed by a behavioral choice in the DG and the SPD. The last SWB instrument measures cognitive life satisfaction (Diener et al. 1985 - SWL).

PWB is measured by only three scales. Following KE, I use the Self-Actualization Index (SAI) by Jones and Crandall (1986) and the Scales of Psychological Well-Being (SPWB) by Ryff (1989). More precisely for the latter measure, I use an index (PWBI), as KE do.[10] Additionally, I implement the Social Well-Being Scale (SoWB) by Keyes (1998). Broadly speaking, PWB tries to measure whether an individual is fully functioning and

[9] Appendix A.3 contains all items upon which these measures are based.

[10] Ryff's Scales of Psychological Well-Being consist of six separate scales. In the abbreviated version that I use, each scale is only measured by tree items. As KE outline, these items were rather chosen for conceptual breadth than for reliability. In order to increase reliability, one can construct an index of SPWB by choosing only the item with the highest average inter-item correlation per scale. For more details see KE's footnote 23.

living a life with purpose and meaning as outlined in Section 2.2. KE's measures of PWB (SPWB and SAI) mainly focus on psychological functioning as a private phenomenon. The SoWB, however, directly tries to evaluate to what extent an individual is flourishing in a society. I extended KE's measure of PWB[11] by SoWB because I think that the social dimension of life is a very important dimension of a deeply satisfying life and that individuals are naturally embedded in social structures and communities.[12]

The last part of the well-being questionnaire is the Marlowe-Crowne (MC) social desirability scale, which is used as a control for social desirability. This scale is especially important in my setting because unlike KE I only use a *single-blind* but not a *double-blind* procedure. In a double-blind procedure, neither the other subjects nor the experimenter can identify individual answers and decisions. With six different games played, such a procedure is, however, not reasonably accomplishable. This means that in my design the experimenter is potentially able to uncover subjects' decisions and answers to well-being questions, which might bias subjects' behavior. The MC social desirability scale is an attempt to detect such a bias.

In the concluding questionnaire, subjects had to answer demographic questions (in-

[11]The question remains whether it is reasonable to include SoWB as a measure of PWB from a psychological view point. In the psychological literature, the term PWB is often used more restrictively in line with Ryff (1989). It has, however, also been used in a much broader meaning just following the general idea of a fully functioning person (see Brown and Ryan 2003) . Someone who understands PWB only in terms of Ryff (1989) may just call my well-being measures eudaimonic well-being instead of psychological well-being.

[12]For the PWB measures, we may be concerned about that the well-being questions asks something that is directly related to behavior in my games, especially for SoWB as a measure of the social component of PWB. So asking whether you trust others and observing people's behavior in a trust game might not be a good idea. SoWB consists of five dimensions, for which two are potentially problematic: *social contribution* and *social acceptance*. The *social contribution* dimension (e.g. Question 4: "I have something valuable to give to the world" - Appendix A.3) is potentially problematic, although we may argue that giving something valuable to the world is not the same as giving something in a dictator game: Entrepreneurs e.g. or successful athletes that behave purely selfish with respect to their fellow citizens may still claim that they contribute something valuable to society, namely jobs in case of the entrepreneur or success in international tournaments and national pride in case of the athlete. Nonetheless, excluding the *social contribution* dimension from SoWB does not qualitatively change my main results (Tables 2.2, A.3, 2.3, 2.7, A.1) for trusting and cooperating in the SPD. It slightly improves results for giving in the DG and it slightly worsens results for punishing in the DG-P (Table 2.2, A.3, 2.3). Overall, excluding this dimension seems to have no systematic effect. Additionally excluding question 13 ("I do not feel responsible to help anybody.") from the SAI does not change my main results in Table 2.2, 2.7, A.1. It weakens results in Table A.3a-A.3b and Table 2.3. Importantly however, there is no change in Table A.3c, which includes the full set of controls and all important differences for the SAI in Table 2.3 remain at least marginally significant. Finally, for the trust decision we might be concerned about questions hinting to what extent people consider others as trustworthy, which is not crucial in other settings because there decisions do not depend on expectations of others. Excluding the *social acceptance* dimension (e.g. Question 14: "I believe that other people are kind." - Appendix A.3) from the SoWB, question 3 ("I believe that people are essentially good and can be trusted") from the SAI, and question 16 ("I have not experienced many warm and trusting relationships with others") from the PWBI weakens results in Table 2.3 for trusting, but all results remain significant at least at the marginal level. In Table A.1 there is no change for the most virtuous and the more virtuous dummy whereas the less virtuous dummy is only marginally significant. More importantly however, there is no qualitative change in Table 2.2, A.3, 2.7. So overall, excluding some questions from PWB measures sometimes has an effect but results remain reasonable robust to conclude that my results do not depend on correlations between single questions and behavior in games.

Table 2.1: Summary of Games

Games	Label	Measure for ...
Dictator game	DG	generosity, altruism
Sequential prisoner's dilemma	SPD	trust and cooperation
Mini-ultimatum game	Mini-UG	second-party punishment
Joy-of-destruction game	JOD-G	spite
SPD with punishment	SPD-P	third-party punishment (low cost)
DG with punishment	DG-P	third-party punishment (high cost)

cluding items about income and expenditures) and to what extent they understood the experiment. Especially, subjects could indicate how sure they were on 1-to-9 scale that they succeeded in answering the well-being questionnaire truthfully. Additionally, I collected a measure of cognitive ability: the Cognitive Reflection Test (CRT) introduced by Frederick (2005). This quick test only consists of three questions. It does not measure cognitive ability per se, but distinguishes quick, impulsive decision makers from more reflective decision makers. Every question of the CRT has an intuitive answer that is incorrect. Although very short, Frederick shows that the Cognitive Reflection Test relates well to more complex measures of cognitive ability. Finally, a ten-item variant of the Big Five Inventory was asked. The Big Five Inventory measures the five most relevant traits of subjects' personality (Gossling et al. 2003).[13]

2.3.2 The Experimental Games

After completing the well-being questionnaire, subjects played six different games summarized in Table 2.1. The following four games are used: a dictator game (henceforth DG), a sequential prisoner's dilemma (SPD), a joy-of-destruction game (JOD-G), and a mini-ultimatum game (Mini-UG). Games five (SPD-P) and six (DG-P) are a variant of the first two games with an additional punishment stage. All games were chosen in order to extend the analysis of KE from generosity (measured by the dictator game) to economically more important behavioral patterns such as trust, (conditional) cooperation, and punishment (positive and negative reciprocity). For all games, subjects made decisions for all roles (role reversal) and were not informed about results until the very end of the experiment.

[13]For all measures, I used German versions when available and carefully translated measures when not: For Bradburn's Positive Affect (PA) and Negative Affect (NA) items, I used the translation by Becker (1982); for the Positive Affect (PAS) and Negative Affect (NAS) Schedules, I used the translation by Krohne et al. (1996); for the Satisfaction with Life Scale (SWL), I used the translation by Schumacher et al., eds (2003); for the Marlowe-Crown Social Desirability Scale (MC), I used the translation by Lück and Timaeus (1969); for the Cognitive Reflection Test (CRT), I used the translation by Oechssler et al. (2009); for the Big Five Inventory, I used the translation by Muck et al. (2007). All other measures were translated by myself.

My DG is a standard variant of the game. The dictator has an endowment of 20 € and can send 0 €, 2 €, ... , 20 € to the recipient who has an initial endowment of 0 € and who has to accept any choice the dictator makes. Unlike in KE's experimental design, each subject makes a decision in the role of the dictator. Additionally, I do not use a double-blind procedure (like KE) because of the five additional games for which implementing double-blindness would have been overall difficult. The DG is implemented to measure generosity (altruism) and to replicate the results of KE.

For the SPD (the game may also be called a bilateral trust game), I use a version proposed by Burks et al. (2010). One player moves first, the other player second. Both player are initially endowed with 10 €. The first mover can only make a binary decision: choose an amount $s_1 \in \{0 €, 10 €\}$ to send to the second mover. The second mover observes this action and chooses an amount $s_2 \in \{0 €, 2 €, 4 €, 6 €, 8 €, 10 €\}$ to send back to the first mover. Any amount sent by the first or the second mover is doubled by the experimenter. This gives the following payoffs: $\pi_i = 10 € - s_i + 2 * s_j$, for $i, j \in \{1, 2\}$ and $i \neq j$. If both players send nothing, both will receive 10 € as a final payoff. If both players send 10 €, both will end up with 20 €. In this situation however, the second mover has an incentive to defect. If the first mover sends 10 € and the second mover sends nothing, the first mover gets nothing and the second mover gets 30 €. According to the subgame perfect equilibrium (with common knowledge of rationality and selfishness), the first mover should anticipate this behavior and send nothing.

Decisions in the SPD measure trust and cooperation (positive reciprocity). I will classify a first mover sending his endowment as a person who trusts. In my experiment, second movers have to make two decisions, one for the case that the first mover sends money and one for the case that the first mover does not send money. Burks et al. (2010) propose the following classification for the second mover: There are three "pure" types: Second movers who always return 0 €, independent of the first mover's choice, are classified as pure *free-riders*. Second movers who choose the most cooperative action available (always return 10 €) are classified as pure *unconditional cooperators*. Second movers who exactly return what has been send to them by the first mover (0 € and 10 €) are classified as pure *conditional cooperators*. For subjects not behaving as one of the pure types, the Euclidean distance between his or her decision and the decision of each of the pure types is calculated. The subject is then assigned to the least distant type category. In my analysis, I follow this classification of Burks et al. (2010) but pool *conditional* and *unconditional cooperators* as *cooperators*, because I observe only very few *unconditional cooperators* (for more details see Appendix A.2.1). The reader may be bear in mind that the overwhelming majority of *cooperators* are *conditional cooperators*. After the SPD, subjects are asked to guess how much money other people return on average when the first player sends money. These non-incentivized believes serve as an estimate for subjects' expectations regarding trustworthiness.

My variant of the JOD-G is adapted from Abbink and Sadrieh (2009) and Abbink and Herrmann (2011). The game structure is the following: Two players have the same endowment and both players make a simultaneous decision to reduce the other player's payoff. Importantly, however, these reduction choices are only implemented with a certain probability. In all other cases, nature reduces both players' payoffs. This means that subjects can burn money and hide behind nature. The JOD-G measures spiteful preferences. Importantly, I will exclude the JOD-G from my analysis because only 5% of subjects showed spiteful preferences which is considerable less than the 26% in the benchmark study of Abbink and Herrmann (2011). Playing the DG and the SPD before the JOD-G, might explain the very low level of spiteful preferences: Subjects might have guessed that the experiment is about cooperative behavior.

The Mini-UG is adapted from Bolton et al. (2005) and Falk et al. (2003). It has a sequential two-stage structure. In the first stage, the proposer has to make a proposal on how to divide a pie of 20 €. However, the proposer is restricted to two choices: In the first proposal, the proposer gets 18 € and the responder gets 2 €. The second proposal is the equal split. In the second stage, the responder can either accept or reject the proposer's proposal. If the responder accepts, the proposal is implemented. In case of rejection, both players get 0 €. In my experiment, subjects make a decision in both roles, and in the role of the responder two decisions are made conditional on the proposer's choice. I am mainly interested in the responder's decision for the case that the proposer chose the unequal split. A subject who rejects such an unequal split will be classified as a person punishing unfair behavior (in a SP-punishment setting).

The SPD-P is adapted from Hoff et al. (2011). Its basic structure is similar to my SPD. However, a third party is added that has the opportunity to punish the behavior of the second mover. Figure 2.2 shows the game tree of the SPD-P and the players' payoffs. The first mover can either send money or keep it. The second mover then has the same binary choice between sending money back or keeping it. Afterwards, the third mover has the opportunity to punish the second mover. There are three crucial aspects about this punishment opportunity: First, I implement punishment as TP-punishment, this means an uninvolved third party has the opportunity to punish. Second, punishment is relatively cheap: The second mover only has to invest 1 € to reduce the second mover's payoff by 5 €. Third, the design allows for both altruistic and antisocial punishment: The second mover can punish both a free-riding and a cooperating second mover. Subjects make decisions in all roles, and my interest is in the punishment decision of the third player. I classify a subject who punishes defectors as an individual punishing unfair behavior in a TP-punishment setting with fairly low costs. A subject punishing cooperators would be classified as an antisocial punisher.

The DG-P modifies my DG and implements a TP-punishment opportunity. Initially, the first player has 15 €, the second player has 5 € and the third player has 20 €. The first

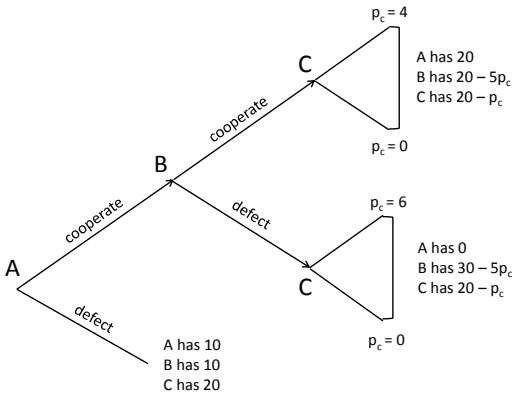


Figure 2.2: Sequential Prisoner's Dilemma with TP-Punishment

player then can send either 0€, 2.5€, or 5€ to the second player. Whereas the second player has no choice to make, the third player then has the opportunity to punish the first player, either with 1€, 2€, 3€, 4€, or 5€. The crucial aspect about this punishment opportunity is that it is rather costly: Investing 1€ in punishment only reduces the first players payoff by 1€. Subjects make decisions in all roles. The DG-P measures (indirect) negative reciprocity and a subject punishing first players who do not share equally[14] will be classified as an individual punishing unfair behavior in a TP-punishment setting with high costs.[15]

To sum it up, six games are implemented. The DG is implemented in order to replicate KE's results. Decisions in the SPD measure "positive" forms of virtuous behavior: trust and cooperation. Decisions in the Mini-UG, the SPD-P, and the DG-P measure "negative" forms of virtuous behavior: punishment. The last three games differ in the following way: Punishment in the Mini-UG is SP-punishment whereas punishment in the SPD-P and DG-P is TP-punishment. In the SPD-P punishment is relatively cheap whereas it is relatively expensive in the DG-P. Finally, it is important to note that only for the decision to trust, virtuous behavior might be instrumentally motivated by earning more money. In all other settings (giving, cooperating and punishing) virtuous behavior is necessarily associated with foregoing own monetary payoff. At least in these settings, subjects should not be guided by (self-concerned) instrumental motivations.

2.3.3 Experimental Procedures

The experiments were conducted at the University of Mannheim in Spring 2012. I run seven sessions with 8 to 18 subjects in each session. In total, 102 subjects participated.

[14] I focus on subjects punishing dictators not sharing equally (instead of those punishing dictators who gave nothing at all) because those subjects are the strictest ones in following the social norm of punishing unfair behavior.

[15] Because of the high costs of punishment, payoff-differences between the first and the third player cannot be increased by punishing. Hence, antisocial punishment should not be expected.

The experimental software was developed in z-Tree (Fischbacher, 2007). For recruitment, ORSEE was used (Greiner, 2004). Out of many possible sequences, the following two were implemented: DG, SPD, JOY-G, Mini-UG, SPD-P, DG-P and SPD, DG, JOY-D, Mini-UG, SPD-P, DG-P. I chose these sequences in order to analyze whether it makes a difference if the DG or the SPD is played first. In general, I did not detect any significant differences in subjects' behavior in the games and hence I will pool the data for most of the further analysis. I will only make use of the non-pooled data when analyzing the change in short-run happiness or mood due to a decision in the DG or the SPD (Appendix A.2.4). For payment, one out of the six games was randomly selected and roles in the selected game were also randomly determined. Additionally, participants received 6 € for completing the well-being questionnaire and the payment resulting from their lottery choice. Sessions lasted about 90 minutes and the average earnings were 24 €.

2.3.4 Experimental Hypotheses

For the case that I indeed find favorable correlations between well-being measures and virtuous behavior, this section proposes different explanations about the underlying causality. The experimental hypotheses are very similar to those of KE (Sec. 3.2). They are only modified to the extent that they suit not only the dictator game setting but also the other games. Furthermore, an additional hypothesis regarding cognitive ability is proposed. Figure 2.3 summarizes the five different hypotheses and lists the variables used to test them. First, I will outline these hypotheses. Afterwards, I will provide some additional remarks regarding my punishment settings.

Virtuousness Hypothesis[16]

One explanation for a favorable correlation between well-being and virtuous behavior is that virtuous behavior causes well-being. However, not in the sense that repeated acts of virtuous behavior increase the stock of PWB but that people behave virtuously because they immediately feel better. This means that virtuous behavior directly increases short-run happiness/SWB.[17] For practical reasons, I can only test this hypothesis for the first two games: This means for trust and cooperation measured by the SPD (and as a replication of KE for giving measured by the DG).[18] If this hypothesis is right, subjects who trust and cooperate (give) in the SPD (DG) should report an improvement in Now Happiness (NH) and the Mood Index (MI) directly after the game, compared to the

[16] It is important not to confuse the Virtuousness Hypothesis with the Virtue Ethics Hypothesis (VEH). The first one refers to changes in *short-run* well-being/ mood whereas the last one refers to *long-run* well-being.

[17] Based on the remarks of Section 2.2, the term *happiness* is used in connection with the hedonic approach and its measure SWB but not with the eudaimonic approach and its measure PWB.

[18] Asking short-run happiness questions more than four times might have reduced the quality of answers because subjects might have got bored by the questions.

measurement before the game. More precisely, subjects who trust and cooperate should score higher in the Now Happiness Difference (NHD) and in the Mood Index Difference (MID).[19] This hypothesis is consistent with warm-glow explanations of giving (Andreoni, 1989; Andreoni, 1990).[20]

Mood Hypothesis

This hypothesis reverses the causality of the first explanation. "People act on emotions" (KE, p. 14) and those who feel good behave more virtuously. Therefore, the Mood hypothesis claims that subjects reporting a higher mood/short-run happiness (MI and NH) just before they make a decision should be more likely to behave virtuously.[21]

Material Well-Being Hypothesis

The next hypothesis claims that both hedonic happiness (long-run SWB) and virtuous behavior are caused by a third factor, namely material well-being. The material well-being (MWB) hypothesis claims that greater MWB leads to higher long-run SWB. Assuming that virtuous behavior is a normal good, higher MWB should also lead to a higher probability of virtuous behavior and additionally stronger virtuous behavior (higher gifts, larger amount sent in the SPD).

Cognitive Ability Hypothesis

This hypothesis claims that another third factor causes hedonic happiness (long-run SWB) and virtuous behavior: namely cognitive ability as measured by the Cognitive Reflection Test (CRT). The cognitive ability hypothesis claims that higher cognitive ability leads to higher long-run SWB and to more virtuous behavior. The underlying argument of this hypothesis is that more intelligent people have greater abilities to cope with life circumstances and might hence be happier. Additionally, higher cognitive ability may lead to greater maturity (potentially in line with the PWB hypothesis) and hence a higher probability of virtuous behavior.

[19]Importantly however, a possible selection bias might occur: Virtuous subjects might gain from behaving virtuously whereas non-virtuous subjects might gain from behaving non-virtuously. And hence, even if we observe no difference between virtuously and non-virtuously behaving subjects, behaving virtuously might still increase happiness for some, namely those who choose to do so. KE run a control treatment without the dictator game decision to account for this problem. I do not implement such a control treatment. Importantly, I observe a difference between virtuously and non-virtuously behaving subjects, at least in the SPD.

[20]KE (p.14) also note that this hypothesis might be consistent with correlations between cooperation/trust and giving and long-run SWB. This is the case if we assume that current cooperation/giving is representative of past patterns of such behavior and that happiness benefits accumulate to improve long-run SWB. However, only evidence about Now Happiness and the Mood Index can be seen as specific evidence for this hypothesis.

[21]For the later analysis, I will use those NH and MI-answers given most closely to the relevant game. KE call this hypothesis *Happiness Hypothesis* instead of *Mood Hypothesis*.

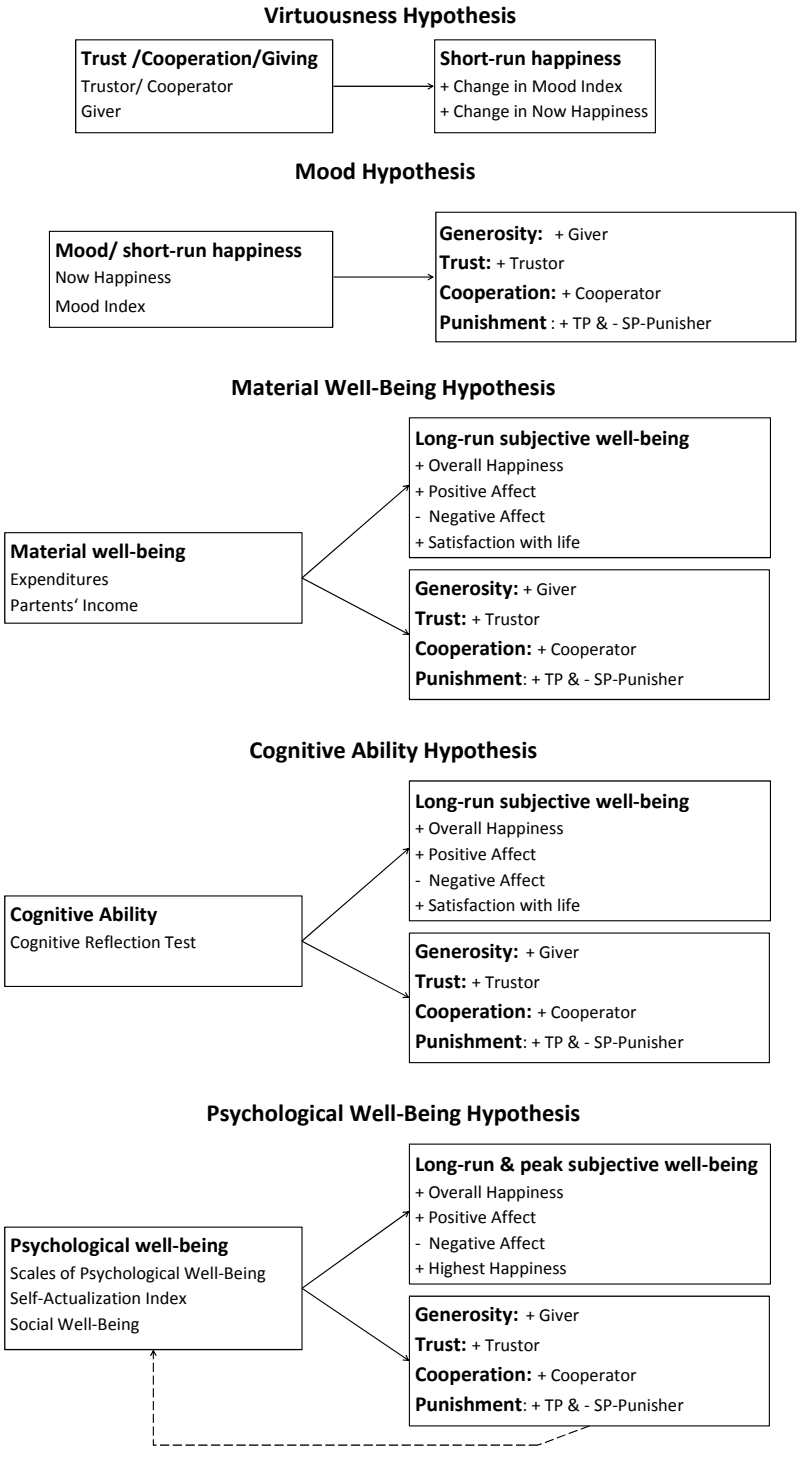


Figure 2.3: Summary of Hypotheses (Compare KE - Fig. 1)

Psychological Well-Being Hypothesis

My main hypothesis is the eudaimonic interpretation of the Virtue Ethics Hypothesis (VEH) described in Section 2.2. Following KE, I name this hypothesis after the measure of eudaimonic well-being: *PWB*. Because my study is not a longitudinal study, I focus on the short-run implication of the model: Subjects enter the experiment with their fixed "stock" of PWB and a high stock of PWB ensures enough psychological resources to generate a high level of hedonic happiness (SWB) and a high level of investment in the stock (virtuous behavior). Therefore, we have to examine whether subjects with higher PWB also report greater long-run subjective well-being (that is higher overall happiness, higher positive affect, lower negative affect and greater highest happiness[22]) and whether more of these subjects behave virtuously.[23]

Additionally, I will also check a slightly modified version of the PWB Hypothesis as a robustness check that reverses the causality between PWB and SWB. The model outlined in Section 2.2 could be wrong. Not PWB, but long-run SWB could cause virtuous behavior (and PWB) although there are no theoretical arguments why this should be the case. I will call this hypothesis the subjective well-being (SWB) hypothesis. Both the PWB hypothesis and the SWB hypothesis should result in similar correlations between PWB, virtuous behavior, and SWB. However, a stronger relationship between virtuous behavior and SWB (PWB) would be evidence in favor of the SWB (PWB) hypothesis.

Frequency of Virtuous Behavior

Another important point outlined by the stock-flow model of Section 2.2 is that the frequency of a kind act and not its size matters. So the PWB hypothesis does not necessarily claim that people with a high level of PWB show stronger virtuous behavior (e.g. larger gifts), as e.g. the MWB hypothesis does, but it claims that these people should behave virtuously in more settings. Nonetheless, also other hypotheses may claim that there is a connection with a high frequency of virtuous behavior: If virtuous behavior increases people's mood, we should expect that those people behaving virtuously more frequently should experience a higher mood increase. If a good mood leads to virtuous behavior, we should expect that the better the people's mood the more frequent virtuous behavior is. Additionally, richer or smarter people could show more frequent virtuous behavior.

[22]Highest happiness is included in the list following KE. The idea behind this is that people striving for self-fulfillment should experience higher peaks in happiness.

[23]As outlined in Section 2, my experiment does not experimentally vary long-run well-being. Hence, it cannot be excluded that the line of causality indicated by the dotted line in Figure 2.3 (repeated acts of virtuous behavior causes PWB) really underlies my results.

Third-Party (TP) versus Second-Party (SP) Punishment

All hypotheses suggest some kind of relationship between well-being and virtuous behavior. But can punishment of unfair behavior and hence harming others be seen as virtuous? Probably only to the extent that punishment of unfair behavior could be an important requirement to maintain high levels of cooperation and trust in a society. In order to evaluate this, three different punishment settings are used: In the Mini-UG, rejection or punishment of unfair offers is done by an involved second party. In the other two games (SPD-P and DG-P), punishment is done by an uninvolved third party. Additionally, punishment costs are high in DG-P and low in SPD-P.

For second-party (SP) punishment in the Mini-UG, at least two interrelated behavioral forces might be relevant in motivating punishment. First, subjects' behavior might be self-centered in the sense that subjects are fully captured by their negative emotions caused by unfair offers. They are angry and retaliate unfair offers by rejecting them and causing a severe monetary loss for the proposer. Second, subjects might be less self-centered and to a smaller extent driven by their emotions but feel (morally) obliged to punish violations of fairness norms, even at own monetary costs. Neuroeconomic findings support the idea that emotions play a key role in rejecting unfair offers in the ultimatum game (see e.g. Sanfey et al., 2003; van 't Wout et al., 2006).[24]

In contrast to the Mini-UG, the punishing subject is not directly affected by unfair behavior in the third-party (TP) punishment settings. Hence, the aspect of retaliation should not play a too important role. Instead, I hypothesize that punishing motivations are not self-centered but targeted towards others. Subjects may feel obliged to punish violations of fairness norms and in addition punishing a defector can be interpreted as an act of kindness towards the negatively affected second party that is unable to punish by itself.

Hence, my conjecture regarding the relationship between punishment and well-being is the following: In the TP-punishment setting, punishment of unfair behavior can be seen as virtuous behavior such as giving, trusting, and cooperating. Hence, there should be a positive relationship between well-being and TP-punishment. SP-punishment, however, is more an act of retaliation or non-virtuous behavior. Hence, I conjecture a negative relationship between SP-punishment and well-being.[25] The difference in punishment costs (DG-P: high/ SPD-P: low) should enable me to discriminate between those subjects who

[24]In my Mini-UG, the strategy method is used. Therefore, I could not ask subjects for their emotions right after receiving an unfair offer. Nonetheless, even in my setting the emotional state directly before making the responder's decision may still be very relevant. Hence, I can still examine whether subjects' short-run or long-run affective SWB has an influence on the decision in the Mini-UG.

[25]For the MWB hypothesis, a positive relation between MWB and TP-punishment hence assumes that TP-punishment is a normal good whereas SP-punishment is assumed to be an inferior good. The motivation behind this distinction is the following: If you are rich, you may well afford to perform kind acts (TP-punishment), but you may also not have to punish others people if you get less money than expected (SP-punishment) just because you already have enough money.

only comply to social norms when complying is fairly cheap and those subjects who have fully internalized the norm. Hence, the positive relationship between TP-punishment and well-being should be more pronounced when punishing is fairly costly.

2.4 Results and Analysis

In this section, I first analyze how well-being and virtuous behavior are related (Section 2.4.1). Afterwards, I examine the five different hypotheses regarding the causal relationship between well-being and virtuous behavior (Section 2.4.2), and finally I check to what extent more frequent virtuous behavior leads to higher well-being (Section 2.4.3).

2.4.1 Well-Being and Virtuous Behavior

Only when the instruments work fine, I can be sure that my analysis of the relationship between well-being and virtuous behavior is relevant and that the questionnaire did not by itself influence people's decisions in the games. Besides the JOD-G, all my other games replicate the conventional results of the literature. Additionally, well-being measures show reasonable correlations with each other, with material well-being (MWB), with cognitive ability (CRT), and with the Marlowe-Crowne social desirability scale (MC). Actually, correlations with the MC scale seem to be higher than in KE's analysis (at least for PWB measures). This potentially indicates that social desirability plays a more important role in a single blind than in a double blind setting. For more details about this analysis, the reader is referred to Appendix A.2.1.

Before analyzing my hypotheses in detail, I will start with some results providing an overview of the relationship between virtuous behavior and well-being. Tables 2.2a and 2.2b provide a first summary of this relationship. Table 2.2a provides the mean scores on well-being measures of those who behaved virtuously in the DG and in SPD and those who did not. For the DG, I compare the mean score on well-being measures of subjects who gave nothing (Nongivers) with those who gave something (Givers). For the SPD, I first compare subjects who did not send any money in the role of the first mover (Nontrustors) with those who did send their endowment (Trustors). Then, I compare the mean score on well-being measures of those classified as Free-riders with those classified as Cooperators. Table 2.2a-2.2b also provide the p-values of one-tail t-tests whether the difference across groups is significant. It is important to bear in mind that well-being data often has a very small variance and hence comparing mean scores might underestimate the magnitude of the differences. Therefore, the column *% SD* indicates how big the difference between those who show virtuous behavior and those who do not is in percent of the standard deviation. For statistically significant results, mean differences are equal to 27% to 75% of a standard deviation, meaning that differences are non-negligible. KE (p. 19) offer some

example to what such differences are comparable. Among others, being unemployed or not, or being one of Forbes superrich leads to comparable differences in well-being measures.

The comparison of Givers and Nongivers is basically a replication of KE (in a single-blind setting). KE's main result is that there is a strong connection between psychological well-being (PWB) and giving and a weaker connection between subjective well-being (SWB) and giving. My results replicate this basic pattern, the connection of well-being measures and giving seems, however, to be overall weaker in my data, potentially due to my single-blind procedure. For SWB, only the Positive Affect Schedule (PAS) and Highest Happiness (HH) are (marginally) significantly higher for Givers than for Nongivers.[26] Hence, there is only fairly weak evidence for a positive relationship between giving and (at least some aspects of) long-run SWB. For mood or short-run happiness (NH, MI), Givers and Nongivers do not have statistically different mean values. Givers and Nongivers also did not experience a statistically distinguishable change in short-run happiness/ mood due to giving or not giving, as indicated by the Now Happiness Difference (NHD) and the Mood Index Difference (MID). The most cognitive-evaluative component of SWB, Satisfaction with Life (SWL), also does not (statistically) differ between Givers and Nongivers. PWB however is greater for Givers for all three measures. All these results are in line with KE's finding. Unlike KE, I find, however, that Givers also have higher material well-being (higher monthly expenditures and higher parent's income) than Nongivers with marginal significance. Additionally, Givers performed better in the Cognitive Reflection Test (CRT). However, these last results are at least not robust across all games, as will become clear below.

Do the basic patterns observed for giving in the DG extend to trusting and cooperating in the SPD? In the SPD, I again observe a weak connection between (long-run) SWB and trust/cooperation and a stronger connection between PWB and trust/cooperation: Subjects who trusted scored higher on Overall Happiness (HH), the Positive Affect Schedule (PAS), and they scored lower on the Negative Affect Schedule (NAS), as expected. Again the other SWB variables do not provide evidence in favor of the reversed relationship between trusting and well-being. For PWB, all three measures indicate that Trustors experienced significantly higher well-being. For the second mover's decision, Cooperators only scored higher on Bradburn's Positive Affect (as a measure of long-run SWB) but for two out of three measures of PWB. Importantly, there is, however, a crucial difference compared to the DG, Free-riders and Nontrustors seem to have experienced a decrease in their mood (MID) (and Free-riders also in their Now Happiness - NHD) after their

[26] A noteworthy aspect is, however, that the different scales measure different aspects of long-run SWB (OH: cognition vs. PA/PAS: positive affect vs. NA/NAS: negative affect) or they measure the same aspect in a different way (e.g. PA vs. PAS). If for all measures the difference between Givers and Nongivers were significant, this would indicate a fairly strong relationship between long-run SWB and giving. If however, only a few scales are significant and the other scales are just insignificant (and not significant for the reversed relationship), this can still be counted as evidence in favor of a weak relationship between some aspects of SWB and virtuous behavior.

Table 2.2a: Well-being for the Dictator Game (DG) and the Sequential Prisoner's Dilemma (SPD) - Mean Values

Well-being measures	**1) DG**				**2) SPD - first move**				**3) SPD - second move**			
	Nongivers	**Givers**	p-value	% SD	**Nontrustors**	**Trustors**	p-value	% SD	**Free-riders**	**Cooperators**	p-value	% SD
Subjective well-bing												
Long-run happiness												
Overall Happiness (OH)	6.21	6.44	0.15	21%	6.05	6.52	**0.02****	41%	6.39	6.34	0.57	4%
Bradburn's Positive Affect (PA)	3.49	3.51	0.47	2%	3.41	3.55	0.29	12%	3.09	3.62	**0.04****	42%
Bradburn's Negative Affect (NA)	1.72	2.03	0.87	23%	1.78	1.98	0.76	15%	1.52	2.03	0.94	37%
Positive Affect Schedule (PAS)	32.54	34.44	**0.06***	32%	32.46	34.43	**0.05***	33%	32.70	34.01	0.17	22%
Negative Affect Schedule (NAS)	21.69	22.14	0.61	6%	23.76	20.95	**0.04****	37%	22.09	21.94	0.46	2%
Highest/ lowest happiness												
Highest Happiness (HH)	7.26	7.59	**0.08***	29%	7.27	7.57	0.10	26%	7.35	7.49	0.29	13%
Lowest Happiness (LH)	3.67	3.41	0.80	17%	3.35	3.60	0.20	17%	3.91	3.39	0.93	36%
Short-run happiness												
Now Happiness (NH)	5.41	5.59	0.25	13%	5.59	5.65	0.43	4%	5.61	5.63	0.47	2%
Mood Index (MI)	39.41	40.63	0.29	11%	41.43	39.66	0.79	16%	41.83	39.86	0.77	18%
Now Happiness Difference (NHD)	0.00	0.10	0.35	8%	-0.16	0.06	0.16	20%	-0.43	0.10	**0.02****	48%
Mood Index Difference (MID)	-1.26	-0.94	0.41	4%	-3.00	0.37	**0.02****	42%	-5.48	0.49	**0.00****	75%
Life satisfaction												
Satisfaction With Life (SWL)	23.92	24.41	0.33	9%	23.89	24.42	0.32	10%	24.57	24.13	0.63	8%
Psychological well-being												
PWB Index (PWBI)	27.46	28.67	**0.07***	31%	27.16	28.80	**0.02****	42%	28.00	28.27	0.39	7%
Self-Actualization Index (SAI)	41.54	43.98	**0.01****	46%	41.51	43.92	**0.01****	45%	40.78	43.71	**0.01****	55%
Social Well-Being (SoWB)	58.03	61.98	**0.01****	44%	56.57	62.69	**0.00****	68%	55.43	61.94	**0.00****	73%
Material well-being												
Monthly expenditures	595.00	688.33	**0.07***	30%	613.51	674.92	0.17	20%	622.17	661.52	0.29	13%
Parents' income	2.85	3.38	**0.06***	31%	3.03	3.26	0.25	14%	3.04	3.22	0.33	10%
Cognitive ability												
Cognitive Reflection Test (CRT)	1.33	1.76	**0.03****	38%	1.59	1.60	0.49	0%	1.30	1.68	**0.08***	34%
N	39	63			37	65			23	79		

Notes: */** indicates significance at the 10%/5% level according to one-tail t-tests of the null hypothesis that subjects behaving virtuously are better off than non-virtuous subjects (e.g. that they have higer positive affect, lower negative effect, higher income, etc.); % *SD* indicates how big the difference between those who show virtuous behavior and those who do not is in percent of the standard deviation; All *p*-values below 10% are in bold type, n=102. Compare KE (Table 3).

non-virtuous behavior. Such a difference does not exist for the two short-run happiness measures, Now Happiness (NH) and the Mood Index (MI). These measures are statistically not different between the two groups, as in the DG. Finally, material well-being (MWB) does not differ across groups whereas cognitive ability (CRT) is higher for cooperators but not for those who trusted.

Do we observe a distinction between different forms of punishment? Table 2.2b presents the same information as Table 2.2a for the Mini-UG, the SPD-P, and the DG-P. The mean comparisons are made between subjects rejecting an unfair offer (Mini-UG) or punishing unfair behavior (Punisher - SPD-P and DG-P) and subjects accepting an unfair offer and not punishing unfair behavior (Nonpunisher). As a reminder, for SP-punishment in the Mini-UG, I conjectured a negative relation with well-being, for TP-punishment, I conjectured a positive relation. The overall finding seems to support this conjecture: Rejecting in the Mini-UG seems to be negatively related to the (affective component) of SWB whereas punishing in the DG-P seems positively related to PWB. So, there seems to be a shift from punishing in a SP-environment to punishing in a TP-environment. For the Mini-UG, rejecting subjects have significantly lower scores in Bradburns' Positive Affect (PA), in the Positive Affect Schedule (PAS), and in Satisfaction with Life (SWL). In addition, they scored higher in the Negative Affect Schedule (NAS) with marginal significance. For PWB, the Index of PWB (PWBI) shows lower values for subjects rejecting unfair offers. For the SPD-P, the result is rather mixed. Only two measures are marginally significant. This result becomes stronger at least for PWB in the DG-P: Two out of three PWB measures are (at least marginally) significant and the last measure (PWBI) has a p-value of 0.10. Additionally, punishing subjects scored higher in the Positive Affect Schedule (PAS). No other variable provides evidence against the expected relationship.

The summary statistics presented so far do not control for demographic variables and personality traits. In order to validate my results, I follow KE (p. 22) and run multivariate regressions, especially ordered logit. Because of space constraints, the details of this analysis are relegated to Appendix A.2.2, which provides a regression analysis with aggregated well-being measures.[27] The main message of these regressions is the following: The major results of the analysis presented so far are robust to including demographic controls and even personality traits. Although regression results do unsurprisingly not confirm the weaker relationship between SWB and virtuous behavior, they do confirm the stronger relationship between PWB and virtuous behavior (across different settings and across different controls). More precisely, the regression analysis confirms that KE's finding of a PWB-giver relationship extents to trusting and cooperating in the SPD and that additionally there seems to be a crucial distinction between SP-and TP-punishment.

[27] Additionally, Appendix A.2.3 provides a complementary regression analysis with non-aggregate well-being measures in order to analyze which specific well-being measures are responsible for the significance of results in the aggregate analysis.

Table 2.2b: Well-being for the Mini-Ultimatum Game (Mini-UG), the Sequential Prisoner's Dilemma (SPD-P) and the Dictator Game (DG-P) with Punishment - Mean Values

Well-being measures	**1) Mini-UG**				**2) SPD-P**				**3) DG-P**			
	Accept	**Reject**	p-value	% SD	**Nonpunisher** (SPD-P)	**Punisher**	p-value	% SD	**Nonpunisher** (DG-P)	**Punisher**	p-value	% SD
Subjective well-bing												
Long-run happiness												
Overall Happiness (OH)	6.35	6.35	0.50	0%	6.36	6.35	0.52	1%	6.35	6.36	0.49	1%
Bradburn's Positive Affect (PA)	3.79	3.24	**0.01****	44%	3.46	3.54	0.38	6%	3.47	3.60	0.33	10%
Bradburn's Negative Affect (NA)	1.81	2.00	0.25	14%	1.76	2.06	0.86	22%	1.95	1.80	0.32	11%
Positive Affect Schedule (PAS)	35.33	32.28	**0.00****	52%	32.84	34.56	**0.07***	29%	33.16	35.44	**0.05****	39%
Negative Affect Schedule (NAS)	20.73	23.07	**0.06***	31%	21.06	22.85	0.88	24%	21.95	22.04	0.52	1%
Highest/ lowest happiness												
Highest Happiness (HH)	7.58	7.35	0.15	20%	7.38	7.54	0.24	14%	7.42	7.60	0.24	16%
Lowest Happiness (LH)	3.46	3.56	0.63	7%	3.72	3.31	0.92	28%	3.51	3.52	0.48	1%
Short-run happiness				0%								
Now Happiness (NH)	5.71	5.44	0.17	19%	5.66	5.48	0.74	13%	5.61	5.44	0.70	12%
Mood Index (MI)	40.10	38.39	0.20	17%	39.48	38.92	0.61	5%	39.30	38.88	0.57	4%
Life satisfaction												
Satisfaction With Life (SWL)	25.00	23.54	**0.09***	27%	24.20	24.25	0.48	1%	24.13	24.52	0.37	7%
Psychological well-being												
PWB Index (PWBI)	28.90	27.59	**0.05****	33%	28.24	28.17	0.53	2%	27.92	29.08	0.10	30%
Self-Actualization Index (SAI)	43.42	42.72	0.26	13%	42.72	43.37	0.27	12%	42.21	45.64	**0.00****	64%
Social Well-Being (SoWB)	61.25	59.78	0.20	16%	59.10	61.79	**0.07***	30%	59.77	62.64	**0.08***	32%
Material well-being												
Monthly expenditures	652.60	652.69	0.50	0%	670.50	635.48	0.71	11%	644.29	678.40	0.32	11%
Parents' income	3.46	2.93	**0.06***	31%	3.30	3.06	0.76	14%	3.16	3.24	0.42	5%
Cognitive ability												
Cognitive Reflection Test (CRT)	1.67	1.54	0.28	12%	1.66	1.54	0.71	11%	1.62	1.52	0.65	9%
N	48	54			50	52			77	25		

Notes: */** indicates significance at the 10%/5% level according to one-tail t-tests of the null hypothesis that subjects behaving virtuously (not rejecting and punishing free-riders) are better off than non-virtuous subjects (e.g. that they have higer positive affect, lower negative effect, higher income, etc.); % *SD* indicates how big the difference between those who show virtuous behavior and those who do not is in percent of the standard deviation; All *p*-values below 10% are in bold type, n=102. Compare KE (Table 3).

Regarding SWB, regression results mainly only confirm the finding that subjects who cooperated in the SPD experienced an increase in their short-run happiness compared to those subjects who did not cooperate.

Before moving to a more detailed analysis of the hypotheses, I would like to provide some general remarks regarding my statistical analysis. Until now and in the following, a large number of individual tests has been and will be used. Using a significance level of 5 (10) percent, we should of course expect that in 5 (10) out of 100 cases non-significant results falsely become significant. Importantly however, as will become clear in the next section, the real aim of my analysis is to check whether different well-being measures are significant across different games and across different types of statistical analysis. Robust significance across games and types of analysis will however not develop by chance.

2.4.2 Analysis of Hypotheses

Compared to KE, I observe weaker correlations between well-being and virtuous behavior (probably due to my single-blind procedure). Nonetheless, as will become clear in this section, my results still provide evidence for the eudaimonic interpretation of the Virtue Ethics Hypothesis. Hence, I confirm KE's finding of the importance of the PWB hypothesis in explaining behavior. Still, I also observe important differences to KE, e.g. for my first hypothesis:

Virtuousness Hypothesis

This hypothesis states that people behave virtuously in order to feel good. This should on average result in a positive change in the current happiness and in the mood of virtuously behaving subjects, or in other words, subjects behaving virtuously should on average report a higher Now Happiness Difference (NHD) and a higher Mood Index Difference (MID) than those who do not behave virtuously. For the DG, my results replicate KE's findings: As indicated in Table 2.2a, Givers stated a mean happiness and mood improvement that is larger than those of Nongivers, but this difference is not significant. This result is confirmed by the regression analysis of Appendix A.2.2 (Table A.3a-A.3c: change in short-run SWB). For the SPD however, subjects, who trusted or cooperated, experienced a higher happiness and mood improvement than those who did not trust and free-ride. More precisely, their happiness and mood improvement is very small, but those who did not trust or did not cooperate experienced a substantial decrease in their mood. The regression analysis of Appendix A.2.2 (Table A.3a-A.3c) confirms these results for the case of cooperating.

In Appendix A.2.4, a more detailed analysis of the virtuousness hypothesis is provided. The main implications of this analysis support the evidence presented so far. Two plausible reasons why non-virtuously behaving subjects experienced a loss in short-run well-being are discussed. Either not behaving virtuously in *both* SPD decisions makes people unhappy

Table 2.3a: Proportion Tests of Those Who Score High and Low on Different Scales

	Mood Hypothesis		**MWB Hypothesis**		**CRT Hypothesis**	**PWB Hypothesis**		
	NH	MI	MWB1	MWB2	CRT	PWBI	SAI	SoWB
Giver (DG)								
High	59.5%	67.3%	65.6%	64.5%	68.5%	65.5%	69.8%	67.9%
Low	72.2%	56.0%	55.3%	57.5%	54.2%	56.8%	53.1%	54.3%
p-value	0.84	0.12	0.15	0.24	**0.07**	0.19	**0.04**	**0.08**
Trustor (SPD)								
High	63.5%	61.2%	64.1%	59.7%	63.0%	72.4%	71.7%	78.6%
Low	64.0%	66.0%	63.2%	70.0%	64.6%	52.3%	55.1%	45.7%
p-value	0.52	0.69	0.46	0.86	0.57	**0.02**	**0.04**	**0.00**
Cooperator (SPD)								
High	83.6%	82.4%	76.6%	75.8%	83.3%	79.3%	86.8%	89.3%
Low	70.2%	72.5%	78.9%	80.0%	70.8%	75.0%	67.3%	63.0%
p-value	**0.05**	0.12	0.61	0.69	**0.07**	0.30	**0.01**	**0.00**
Reject (Mini-UG)								
High	50.9%	49.0%	54.7%	45.2%	51.9%	46.6%	50.9%	51.8%
Low	55.3%	56.9%	50.0%	65.0%	54.2%	61.4%	55.1%	54.3%
p-value	0.33	0.21	0.68	**0.03**	0.41	**0.07**	0.34	0.40
Punisher (SPD-P)								
High	49.1%	51.0%	50.0%	48.4%	46.3%	51.7%	50.9%	53.6%
Low	53.2%	51.0%	52.6%	55.0%	56.3%	50.0%	51.0%	47.8%
p-value	0.66	0.50	0.60	0.74	0.84	0.43	0.50	0.28
Punisher (DG-P)								
High	27.3%	25.5%	25.0%	22.6%	24.1%	29.3%	35.8%	32.1%
Low	21.3%	23.5%	23.7%	27.5%	25.0%	18.2%	12.2%	15.2%
p-value	0.24	0.41	0.44	0.71	0.54	**0.10**	**0.00**	**0.02**

Notes: For each variable (e.g. NH, MI), subjects are split into those who score at or above the median in terms of that variable (High) and those who are below the median (Low). For every variable the High/Low row shows the percentage of subjects giving, trusting, cooperating etc. in each group. p-values are provided by one-tail z-tests of the hypothesis that the percentage of the High group exceeds the percentage of the Low group. Only for the Mini-UG these one-tailed z-tests test the reverse hypothesis. All p-values below 10% are in bold tye. Compare KE (Table 6).

because they are not using the efficiency enhancing power of the game (sent money is doubled) or free-riding as such has much worse effects than not trusting. In either case, I can conclude that the evidence in favor of the virtuousness hypothesis is mixed. In the SPD, it seems to be important but not in the DG. Hence, the hypothesis' predictive power might only be valid in specific game settings but it should a prior not be neglected.

Mood Hypothesis

The next hypothesis reverses the causality of the former hypothesis: Mood or short-run happiness causes virtuous behavior. More precisely, subjects scoring high in Now Happiness (NH) and in the Mood Index (MI) should give, trust, cooperate, and punish more (as a third party). Summary statistics depicted in Table 2.2a-2.2b provide no support for this hypothesis. The same is true for the regression analysis of Appendix A.2.2 (Table A.3a-A.3c). An additional test is provided in Table 2.3a. Following KE (p. 25), in this table, I split subjects at the median value for NH and MI (and other variables). For every game, I then test whether the proportion of people giving, trusting etc. is higher for those

Table 2.3b: Difference in Means of Those Who Score High and Low on Different Scales

	Mood Hypothesis		**MWB Hypothesis**		**CRT Hypothesis**	**PWB Hypothesis**		
	NH	MI	MWB1	MWB2	CRT	PWBI	SAI	SoWB
Giving (DG)								
High	4.24	5.04	4.66	4.97	4.93	4.76	5.47	4.93
Low	5.89	4.00	4.32	3.85	4.08	4.23	3.51	4.04
p-value	0.93	0.12	0.35	0.10	0.17	0.27	**0.01**	0.16
Cooperation (SPD)								
High	8.36	8.04	7.16	7.00	7.52	7.66	8.30	8.50
Low	6.30	6.78	7.84	8.05	7.29	7.09	6.45	6.09
p-value	**0.01**	**0.08**	0.77	0.88	0.40	0.26	**0.02**	**0.00**
Punihment (SPD-P)								
High	2.22	2.36	2.27	2.20	2.04	2.35	2.45	2.36
Low	2.06	1.93	1.93	2.06	2.27	1.88	1.82	1.89
p-value	0.38	0.19	0.25	0.39	0.68	0.17	**0.10**	0.17
Punishment (DG-P)								
High	0.67	0.69	0.75	0.65	0.80	0.83	1.09	0.89
Low	0.66	0.65	0.53	0.70	0.52	0.45	0.20	0.39
p-value	0.48	0.44	0.21	0.58	0.15	**0.09**	**0.00**	**0.03**

Notes: For each variable (e.g. NH), subjects are split into those who score at or above the median in terms of that variable (High) and those who are below the median (Low). For every variable the High/Low row shows the mean gift/ mean amount sent/ mean amount of punishment for each group. p-values are provided by one-tail t-tests of the hypothesis that the mean value of the High group exceed the mean value of the Low group. All p-values below 10% are in bold type. Compare KE (Table 6).

scoring at or above the median. Only for the Mini-UG this test is reversed. For both variables, NH and MI, in all six setting, there is only one significant result: For high NH subjects, the fraction of cooperators is significantly higher than for low NH subjects. This result is further corroborated by Table 2.3b. Instead of fractions, this table looks at mean values (mean gift, mean amount sent, mean amount of punishment) in those settings in which such an analysis is meaningful. For all four situations, again only the mean amount sent by second movers in the SPD by high NH/MI subjects is significantly higher than the amount sent by low NH/MI subjects. Overall, support for this hypothesis is very weak. Only in one setting (second mover in SPD) and by one type of analysis (median split of well-being measures), I find a significant result.

Material Well-Being Hypothesis

This hypothesis states that material well-being (MWB) is the tertiary causal factor of both virtuous behavior and well-being. Regarding the MWB-virtuousness relationship, Table 2.2a and 2.2b only provide some support for the DG and the Mini-UG: Givers have higher monthly expenditures (MWB1) and state higher parents' income (MWB2). Additionally, subjects rejecting unfair offers report lower parents' income. These results are only partly supported by the analysis of Table 2.3a-2.3b. Subjects with higher material well-being are more likely to give, but the difference to those subjects low in MWB is not significant. Out of 20 tests provided by Table 2.3a-2.3b only one is significant at a

conventional level: Subjects reporting lower parents' income (MWB2) are more likely to reject an unfair offer in the Mini-UG. However, this relationship is not corroborated with subjects' monthly expenditures (MWB1). Overall, there is only very limited support for the MWB-virtuousness relationship.

Regarding the MWB-happiness relationship, Appendix A.2.1 (Table A.2b) provides spearman correlations between well-being (SWB and PWB) and material well-being (MWB). Table A.2b gives a first indication that a weak relationship between MWB and happiness might exist. Focusing on long-run happiness/SWB and life satisfaction, four correlations (out of 18) are significant at the 5% level. Table 2.4 further analyzes this relationship analogus to Table 2.3a: Subjects are split up at the median value of the two material well-being measures and mean values of the relevant hedonic happiness measures (OH, PA, PAS, NAS, and SWL) are reported separately for high and low MWB subjects. Unlike KE, who find no significant difference at all, differences between high and low subjects are (at least weakly) significant in six out of ten cases. For OH, PAS, and SWL differences are significant for both material well-being measures. In general, a positive relationship between income, especially relative income (Easterlin, 1974), and hedonic happiness is not an unusual finding in the literature. A possible reason why I am able to detect such a relationship, whereas KE are not, could be the sample size: KE have only 48 observations compared to 102 observations in my experiment. Additionally, KE only have undergraduate subjects whereas I do not restrict my sample only to undergraduates. In sum, I find no conclusive support that material well-being causes virtuous behavior although a favourable correlation between hedonic well-being and material well-being seems to exist. Because of the missing link between MWB and virtuous behavior, I find no support that MWB causes both virtuous behavior and hedonic well-being, as the MWB hypothesis states.

Cognitive Ability Hypothesis

Is cognitive ability the tertiary causal factor of both virtuous behavior and happiness? Regarding the relationship of cognitive ability and virtuous behavior, Table 2.2a-2.2b only support a positive relationship for giving in the DG and cooperating in the SPD. This relationship is corroborated by Table 2.3a (with mariginal significance). Subjects with high values in the Cognitive Reflection Test (CRT) are more likely to give and to cooperate. However, this group does not provide a significantly higher mean gift or a higher mean amount sent (Table 2.3b). Importantly, the relationship between PWB and virtuous behavior remains significant even after controlling for the CRT, as can be seen in the regression analysis of Appendix A.2.2 (Table A.3a-A.3c). So cognitive ability is not the one tertiary causal factor explaining virtuous behavior. One explanation for the weak relationship between virtuous behavior and cognitive ability could be that cognitive ability

Table 2.4: Results on Hedonic Happiness and Material Well-Being/Cognitive Ability

		Mean scores				
		OH	PA	PAS	NAS	SWL
MWB measure						
Monthly Expenditures	High	6.56	3.42	34.47	21.88	24.88
	Low	6.00	3.63	32.45	22.13	23.13
	p-value	**0.01**	0.79	**0.05**	0.43	**0.06**
Parents' Income	High	6.48	3.63	34.53	21.81	25.55
	Low	6.15	3.30	32.45	22.23	22.18
	p-value	**0.08**	0.10	**0.04**	0.39	**0.00**
Cognitive Ability						
Cognitive Reflection Test	High	6.48	3.65	34.28	21.63	24.65
	Low	6.21	3.33	33.08	22.35	23.75
	p-value	0.12	0.11	0.16	0.32	0.21

Notes: Subjects are split into those who score at or above the median MWB/ CRT and those who are below. For each group mean values of different SWB measures (OH, PA, PAS, NAS, SWL) are reported. Additionally, p-values for a one-tail t-test of the null hypothesis that the High SWB group is better off than the Low group are reported. All p-values below 10% are in bold type. Compare KE (Table 7).

leads to higher maturity in line with the PWB hypothesis.[28] So, the PWB hypothesis provides an explanation on its own why people behave virtuously, but cognitive ability might facilitate making use of this explanation.[29]

Regarding the relationship between cognitive ability and happiness, Table A.2b (Appendix A.2.1) provides no initial support for our hypothesis. This is corroborated by Table 2.4. Subjects scoring higher in the CRT are on average happier but none of these differences is even marginally significant at a conventional level. We might interpret this result in line with the general finding in the literature that IQ and happiness are related but only to a very small extent (Argyle, 1999). With my sample size, I might not able to statistically detect this comparatively weak relationship.

Overall, there is some evidence that cognitive ability and virtuous behavior are moderately related, at least in some settings. However, I do not find statistically robust evidence for the relationship between cognitive ability and happiness. So, I only find some evidence that cognitive ability has an impact on virtuousness (and maybe on well-being). Hence, cognitive ability does not seem to be the major driving force in my experiment but may provide an additional explanation for virtuous behavior, that has not been analyzed by

[28] This argument is at least partially supported by the following data: High CRT subjects score higher on the PWBI (High: 28.8, Low: 27.4) and the SAI (High: 43.6, Low: 42.3) but slightly lower on the SoWB (High: 60.1, Low: 60.9) than low CRT subjects.

[29] An important aspect, however, that I am unable to rule out is that the weak positive relationship between cognitive ability and virtuous behavior may depend on my measure of cognitive ability. As outlined earlier, the Cognitive Reflection Test (CRT) does not measure cognitive ability per se but whether people can avoid intuitive but wrong answers. Although this measure is highly correlated with cognitive ability, it might be that not cognitive ability as such but cognitive impulse control is really related to well-being. If this is the case, not cognitive ability but cognitive impulse control may facilitate making use of the nexus between virtues and well-being.

Table 2.5: Results on Subjective and Psychological Well-Being

		Mean scores				
		OH	PA	PAS	NAS	HH
PWB measure						
Index of PWB	High	6.69	3.83	35.78	20.66	7.69
	Low	5.91	3.07	31.00	23.70	7.16
	p-value	**0.00**	**0.00**	**0.00**	**0.02**	**0.01**
Self-Actualization Index (SAI)	High	6.47	3.66	34.92	19.64	7.53
	Low	6.22	3.33	32.41	24.49	7.39
	p-value	0.14	**0.09**	**0.02**	**0.00**	0.27
Social Well-Being (SoWB)	High	6.64	3.70	35.02	20.89	7.73
	Low	6.00	3.26	32.13	23.28	7.13
	p-value	**0.00**	**0.04**	**0.01**	**0.06**	**0.00**

Notes: Subjects are split into those who score at or above the median of different PWB measures and those who are below. For each group mean values of different SWB measures (OH, PA, PAS, NAS, SWL) are reported. Additionally, *p*-values for a one-tail t-test of the null hypothesis that the High PWB group is better off than the Low group are reported. Note, that being better off for the NAS means having lower values. All p-values below 10% are in bold type. Compare KE (Table 8).

KE.

Psychological Well-Being Hypothesis

Table 2.2a-2.2b (and the regression analysis of Appendix A.2.2 - Table A.3a-A.3c) have shown that there is a fairly robust relationship between virtuous behavior and PWB. Although this relationship seems to be not as strong as in KE's analysis, it remains the statistically most significant and robust relationship in my analysis. Additional evidence comes from Table 2.3a. Do those who score higher in PWB measures have a higher probability to behave virtuously? For giving, trusting, and cooperating, subjects scoring high in any of the three PWB measures have a higher likelihood of behaving virtuously. This difference between high PWB and low PWB subjects is always significant for the SAI and for SoWB. Regarding punishment, I find that for the DG-P the probability of high PWB subjects to be a punisher is significantly higher than for low PWB subjects (for all three measures of PWB). For the Mini-UG and the hypothesized reverse relationship, only the PWBI is (marginally) significant. Regarding the PWB-SWB relationship, Table 2.5 presents results analogous to those of Table 2.3a and 2.4. For all three PWB measures, subjects scoring high in these measures have a higher (average) value for the relevant SWB measures (OH, PA, PAS, NAS, HH). These differences are significant in 13 out 15 cases.

In general, the observed PWB-virtuousness relationship and the PWB-SWB relationship are consistent with the PWB hypothesis. Additionally, these relationships seem to be robust across different ways of analysis (Table 2.2a-2.2b, A.3a-A.3c, 2.3a-2.3b, 2.5) and for different settings of virtuous behavior (giving, trusting, cooperating, and third-party punishment with high costs). As outlined in Section 2.3.4, one might hypothesize that long-run SWB (instead of PWB) causes both virtuousness and PWB (and call this the

SWB hypothesis) and hence explains the PWB-virtuousness-SWB relationship. The general observation that the PWB-virtuousness relationship, however, is fairly robust whereas the SWB-virtuousness relationship seems to be fairly weak speaks in favor of the PWB hypothesis and against the SWB hypothesis. The PWB hypothesis states that PWB directly causes (long-run) SWB and (short-run) virtuous behavior. The indirect relationship between SWB and virtuousness should be much weaker whereas the SWB hypothesis just predicts the opposite. So my data seems to be much more in favor of the PWB hypothesis than of the SWB hypothesis.[30]

Hence overall, my data confirms that KE's finding that the PWB hypothesis best explains giving in the DG extends to economically more important patters of virtuous behavior such as trust, cooperation and punishment of unfair behavior. My analysis, however, also reveals that others factors cannot be totally neglected: In the SPD, changes in short-run happiness due to subjects' decisions play an important role. Additionally, cognitive ability seems to provide a (weak) complementary explanation for virtuous behavior.

Third-Party (TP) versus Second-Party (SP) Punishment

Regarding punishment, the data of Section 2.4.1 supports the idea that there is a crucial distinction between SP- and TP-punishment. Table 2.2a-2.2b suggest that SP-punishment in the Mini-UG is negatively related to long-run affect (SWB) and that TP-punishment in the DG-P is positively related to PWB. This relationship is confirmed by Table 2.3a-2.3b. Only for the DG-P, high PWB subjects have a significantly higher probability to punish than low PWB subjects. Overall, my experimental evidence supports my conjectures about the motivational forces of punishment: In the DG-P, these forces seem to be (positively) related to PWB, punishment can be seen as an act of kindness. In the Mini-UG, these forces seem (at least on average) to be negatively related to the affective component of long-run SWB. This would be in line with findings in the literature that emotions play a major role in rejections in the ultimatum game. On average, punishment in the Mini-UG seems to be an act of retaliation and not an act of kindness. Importantly however, the next section will provide some additional insights regarding individual behavior and punishment.

[30]In line with KE (p. 28), it is possible to provide a more formal test. If the SWB Hypothesis is correct, all the covariance between virtuous behavior and PWB should operate through SWB. If we regress PWB on SWB, adding a dummy for virtuous behavior should not enhance the explanatory power of the regression. Adding e.g. the giver dummy to a regression of SWB on PWB instead, should increase the goodness-of-fit. If on the other hand the PWB Hypothesis is correct, the predictions are reversed: Adding the dummies to a regression of PWB on SWB should increase explanatory power whereas adding the dummies to a regression of SWB on PWB should not. Although not presented here, such an analysis indeed further supports the PWB hypothesis.

2.4.3 Frequency of Virtuous Behavior

Finally, my within-subject design allows me to analyze to what extent subjects show consistent virtuous behavior across games and whether those subjects behaving virtuously more frequently also report higher well-being. The stock-flow model of Section 2.2 suggests that more frequent acts of virtuous behavior lead to a higher stock of PWB and that a higher stock of PWB makes more frequent virtuous behavior more likely. Hence, I expect that those people behaving virtuously in one setting also have a high likelihood of behaving virtuously in another setting. Table 2.6 confirms this expectation. It provides the correlation coefficients (and significance levels) between the relevant decisions made in the experiment. Giving (DG), trusting (SPD), cooperating (SPD), and TP-punishment (SPD-P and DG-P) are clearly correlated with each other. SP-punishment, however, is only correlated with TP-punishment, not with giving, trusting and cooperation. So, results suggest that behaving virtuously in one setting makes virtuous behavior in another setting more likely. Only (non-altruistic) SP-punishment in the Mini-UG is not correlated with all the other decisions. Hence, I again observe a crucial distinction between the different forms of punishment.

Table 2.6: Correlations Between Decisions

	DG	SPD 1	SPD 2	Mini-UG	SPD-P	DG-P
Giving (DG)	1.00					
Trust (SPD 1)	0.35	1.00				
	0.001					
Cooperation (SPD 2)	0.41	0.62	1.00			
	0.000	**0.000**				
Rejection (Mini-UG)	-0.09	-0.01	-0.08	1.00		
	0.445	0.866	0.480			
Punishment (SPD-P)	0.34	0.20	0.21	0.19	1.00	
	0.000	**0.055**	**0.033**	**0.066**		
Punishment (DG-P)	0.31	0.14	0.19	0.19	0.62	1.00
	0.001	0.106	**0.046**	**0.021**	**0.000**	

Notes: p-values below 10% are in bold type; Ordinal variables: Giving, Cooperation, Punishment; Dichotomous variables: Trust, Rejection; Spearman's ρ for two ordinally scaled variables; Pearson's correlation for two dichotomous variables; Rank biserial correlation for one ordinal and one dichotomous variable, n=102.

Additionally, my within-subject design also allows me to look at single individuals in order to further examine the consistency of choices across games. It seems reasonable to focus first on the non-punishment choices in the DG and the SPD because for these games there is no doubt about what can be considered as virtuous behavior. 61 out 102 subjects behave consistently in the following way: 47 subjects give, trust, and cooperate and 14 do not give, do not trust, and do not cooperate. All other subjects behave virtuously in some situations wheres they do not behave virtuously in others. So overall, these figures support the idea that quite a number of subjects are consistently virtuous. Additionally,

it seems reasonable that not all subjects show such kind of strong consistency. We may e.g. conjecture that it is a social norm to cooperate in the SPD whereas trusting might not be considered such a norm or at least only considered to be such a norm to a more limited extent. Free-riders are potentially punished by others whereas those who do not trust may not be punished for their decision. Hence, it is no surprise that more people cooperate than trust and that for this reason not all people show strong consistency in their behavior.

Overall, I observe consistent behavior across games, but do those subjects behaving virtuously more frequently also report higher well-being? In order to analyze this conjecture, I again focus on the behavior in the DG and the SPD, in which virtuous behavior can be clearly distinguished from non-virtuous behavior. Subjects make three decisions. I will call those subjects who behave virtuously in all three decisions *most virtuous* subjects and those who behave virtuously in no decision *least virtuous* subjects. Subjects behaving virtuously in one (two) decisions are called *less virtuous* (*more virtuous*) subjects. Table 2.7[31] presents mean values of well-being measures for all four types in a similar format to Table 2.2a-2.2b: The last column provides the p-value of the non-parametric Jonckheere's trend test (Jonckheere, 1954) that is an extension of the Wilcoxon rank sum test. It tests whether well-being measures increase with the frequency of virtuous behavior.[32]

Table 2.7 mainly supports the findings reported so far. Most importantly, evidence that well-being increases with the frequency of virtuous behavior is strongest for PWB. The Jonckheere's test is significant for all three PWB measures. For long-run SWB, only two measures, the Positive Affect Schedule (PAS) and Overall Happiness (OH), seem to increase with the frequency of virtuous behavior. With marginal significance, the change in short-run happiness, measured by the Now Happiness Difference (NHD) and the Mood Index Difference (MID), increases with the frequency of virtuous behavior. This result is potentially driven by the fact that not behaving virtuously in both SPD settings leads to a decrease in short-run happiness as outlined before. Additionally with marginal significance, the Cognitive Ability Test (CRT) increases with the frequency of virtuous behavior, reflecting the weak evidence in favor of the cognitive ability hypothesis found before. Overall, the different levels of significance of the Jonckheere's trend test seem to mirror the predictive power of the different hypotheses found in Section 2.4.2. There is clear evidence that more frequent virtuous behavior is related to higher well-being, especially eudaimonic well-being, supporting again a eudaimonic interpretation of the Virtue Ethics Hypothesis.

[31] In this table, for short-run happiness, the first request of Now Happiness (NH) and the Mood Index (MI) is used. For the change of short-run happiness, the change of NH and the MI after playing the DG and the SPD are used.

[32] Including additionally the behavior in the DG-P, SPD-P, and even the Mini-UG does not change results very much. When focusing, however, only on the three punishment decision, results do not qualitatively change much only if rejection in the Mini-UG is considered as non-virtuous behavior.

Table 2.7: Well-being According to How Virtuous Subjects Behave - Mean Values

Well-being measures	DG and SPD				
	Least virtuous	**Less virtuous**	**More virtuous**	**Most virtuous**	**Jonckheere's Trend Test** p-value
Subjective well-bing					
Long-run happiness					
Overall Happiness (OH)	6.14	6.19	6.28	6.51	**0.10***
Bradburn's Positive Affect (PA)	3.21	3.19	3.84	3.51	0.24
Positive Affect Schedule (PAS)	31.79	33.25	32.92	34.87	**0.02****
Negative Affect Schedule (NAS)	21.57	22.63	24.12	20.72	0.78
Highest/ lowest happiness					
Highest Happiness (HH)	7.14	7.38	7.40	7.62	0.12
Lowest Happiness (LH)	4.00	3.38	3.24	3.55	0.59
Short-run happiness					
Now Happiness (NH)	5.50	5.44	5.44	5.62	0.27
Mood Index (MI)	43.07	39.56	39.20	42.06	0.35
Now Happiness Difference (NHD)	-0.21	-0.44	0.08	0.26	**0.07***
Mood Index Difference (MID)	-7.29	-2.25	-1.84	-0.23	**0.05***
Life satisfaction					
Satisfaction With Life (SWL)	24.79	23.13	24.04	24.53	0.25
Psychological well-being					
PWB Index (PWBI)	28.14	26.69	27.36	29.19	**0.02****
Self-Actualization Index (SAI)	40.57	41.69	42.24	44.68	**0.00****
Social Well-Being (SoWB)	52.93	58.88	60.96	63.00	**0.00****
Material well-being					
Monthly expenditures	652.50	554.06	603.20	712.55	0.10
Parents' income	2.57	3.44	3.00	3.36	0.18
Cognitive ability					
Cognitive Reflection Test (CRT)	1.21	1.50	1.68	1.70	**0.10***
N	14	16	25	47	

Notes: */** indicates significance at the 10%/5% level according to Jonckheere's trend test of the null hypothesis that subjects in all groups have the same mean value against the alternative hypothesis that subjects with more frequent virtuous behavior have higher mean values; All *p*-values below 10% are in bold type, n=102.

But are these results robust to including controls? Table A.1a-A.1c (Appendix A.1) provide a regression analysis analogous to Table A.3a-A.3c (Appendix A.2.2): For each aggregated well-being measure, however, three dummies are included in each regression (*less virtuous*, *more virtuous* and *most virtuous* subjects). These regressions mainly confirm the main findings of Table 2.7. For PWB, all dummies are significant and the coefficients increase with the frequency of virtuous behavior. The difference between less virtuous and more virtuous subjects is never significant. Differences between less virtuous or more virtuous versus most virtuous subjects are often at least marginally significant.[33] For the change in short-run happiness, at least the *most virtuous* dummy is significant. The two other dummies are insignificant but have the expected sign.

This analysis may also shed light on the difference between SP- and TP-punishment. The composition of those subjects who punish unfair behavior changes meaningfully between SP- and TP-punishment: In the Mini-UG, only 24 out 52 (46%) of rejecting subjects are most virtuous subjects. Hence, a slight majority of rejecting subjects does not show most virtuous behavior in the DG and the SPD. For TP-punishment, however,

[33]Table A.1a - less virtuous vs. most virtuous: Z = 1.60, p = 0.109, more virtuous vs. most virtuous: Z = 1.69, p = 0.090; Table A.1b - less virtuous vs. most virtuous: Z = 1.81, p = 0.071, more virtuous vs. most virtuous: Z = 1.59, p = 0.111; Table A.1c - less virtuous vs. most virtuous: Z= 1.86, p = 0.063, more virtuous vs. most virtuous: Z = 1.88, p = 0.060.

a majority of punishing subjects belongs to the category of the most virtuous subjects: In the SPD-P, 31 out 52 (59%) and 16 out 25 (64%) in the DG-P. On the one hand, this composition effect confirms the finding so far that there is a difference between SP- and TP-punishment.[34] On the other hand, quite a number of most virtues subjects also punish in the Mini-UG. Hence, the conclusion of the last section about different forms of punishment has to be qualified: It might well be that not for all subjects punishment in the Mini-UG is an act of retaliation. Most virtues subjects might punish for non-selfcentered reasons in the Mini-UG as they also do in SPD-P and the DG-P. These subjects are however not the majority and those subjects who behave less virtuously lower the average well-being of all subjects who reject in the Mini-UG. Hence, on average, a distinction between SP- and TP-punishment arises, as seen before.

2.5 Conclusion

Is there a nexus between well-being and virtuous behavior? The answer seems to depend on the approach of well-being one uses. For a hedonic interpretation of well-being, I find only limited evidence for such a nexus and additionally this evidence has not proven to be very robust to including additional controls. For a eudaimonic interpretation however, I find conclusive evidence in favor of a positive relationship between psychological well-being and virtuous behavior. My data is mostly in line with the hypothesis that virtuous behavior is both a long-run cause as well as a short effect of eudaimonic well-being. Hence, KE's main finding of the importance of psychological well-being (as a measure of eudaimonic well-being) for giving in the dictator game is also valid for trust, cooperation and (expensive) third-party punishment. To this extent, there is a nexus between virtuous behavior and eudaimonic well-being, supporting a eudaimonic interpretation of the Virtue Ethics Hypothesis. Hence, well-being and virtuous behavior are connected with each other to the extent that we think about well-being not just in terms of pleasure and pain (as the hedonic well-being approach does) but in terms of whether people have something like a purpose in life or whether they are striving for self-fulfillment. My results hence also at least partially question the primary focus of economists on subjective well-being when working on well-being. In my view, subjective well-being is a very important measure. However, my analysis shows that psychological well-being can be a very useful complement for measuring well-being.

Importantly and in contrast to KE, I also find that other factors besides eudaimonic

[34]For completeness, I note that the difference between punishing in the Mini-UG and in the DG-P does not solely emerge form the described composition effect. Well-being measures (PWB) of the most virtuous subjects punishing in the DG-P are higher than those rejecting in the Mini-UG (Mini-UG - PWBI: 28.9, SAI: 44.2, SoWB: 62.3; DG-P - PWBI: 29.8, SAI: 48.1, SoWB: 65.2). But well-being measures of all other subjects punishing in the Mini-UG (Mini-UG - PWBI: 26.5, SAI: 41.5, SoWB: 57.8) are still lower than those of the most virtuous subjects.

well-being cannot be totally neglected. In some settings, virtuous behavior may directly increase people's mood. Additionally, cognitive ability maybe a third factor that facilitates benefiting from the nexus between virtues and well-being. Regarding punishment of unfair behavior, my analysis suggests that the punishment of a directly involved second party might often be motivated by emotional forces that do not lead to higher well-being. This potentially reflects the biblical demand "to turn the other cheek if someone slaps you on your right cheek. Punishment provided by an uninvolved third party (at high costs) in contrast seems to be driven by motivational forces in line with the eudaimonic idea of an individual striving for human growth and self-fulfillment. Hence, punishment may not only differ to what extent it is socially beneficial or not but to what extent it is personally beneficial in terms of well-being or not and we may conjecture that both ideas could be related. Finally, my within-subject design shows that virtuous behavior in different settings is significantly correlated. Even more important, subjects that behave virtuously more frequently report higher (eudaimonic) well-being. These findings are clearly in line with the Virtue Ethics Hypothesis and hence provide additional support for it.

The nexus between virtues behavior and eudaimonic well-being may also provide us with an explanation why some people behave virtuously and others do not. Subjects' individual insight in the relationship of virtuous behavior and well-being (and their development of a corresponding character trait), or in other words, their ethical maturity may influence whether a person behaves virtuously or not. Mature people know that virtuous behavior increases their well-being whereas less mature people are unaware of this insight. Hence, mature people may be motivated by this insight to develop a suitable character trait and behave virtuously whereas the less mature are not. The weak relationship between cognitive ability and virtuous behavior might be an indication that cognitive ability may facilitate the development of ethical maturity. So, how can the fraction of virtuously behaving citizens in a society be increased? To the extent that we consider behavioral patterns such as altruism, trust, (conditional) cooperation, and punishing of unfair behavior as an important ingredient or even as a precondition of economic prosperity (as outlined in the Introduction), my analysis provides a hint that those factors guiding us to (eudaimonic) well-being may also lead us to a flourishing society, especially if we assume both lines of causality between eudaimonic well-being and virtuous behavior. Future research may determine to what extent the criteria used for psychological well-being can guide us to make people both more happy and more virtuous. Additionally, as already pointed out by KE (p. 1), there are non-material benefits of policies that promote "charitable donations, volunteerism, service education, and [...] community involvement" probably because these policies make people aware of the nexus between virtuous behavior and well-being and foster the development of a corresponding character trait. Not taking these benefits into account, may deteriorate both individual well-being and economic prosperity.

Chapter 3

Do Reference Points Lead to Wage Rigidity? Experimental Evidence[1]

3.1 Introduction

It is a long-standing puzzle in economics why wages show downward rigidity (Bewley 1999; Fehr and Goette 2005). Different theories of wage rigidity[2] have been proposed and in order to examine the validity of different explanations, questionnaires studies interviewing managers have been performed (e.g. Blinder and Choi 1990; Agnell and Lundborg 1995; Campbell and Kamlani 1997; Bewley 1999). The empirical evidence derived from these studies suggests an intuitive answer to the wage rigidity puzzle: Firms do not cut wages because they are constrained by fairness considerations. In labor markets with incomplete contracts, firms have to rely at least partially on workers' intrinsic motivation. Wage cuts, however, may lead to wage levels below workers' perceived fair wage, resulting in lower workers' effort. Firms anticipate this and do not cut wages because they fear a decrease in work morale. The validity of this fear is, however, not obvious. (Outcome-based) fairness models (e.g. Fehr and Schmidt 1999; Bolton and Ockenfels 2000) predict that wages should be cut in recession because otherwise only firms' profits would fall if wages remained constant. Hence in recession, the split of surplus between firms and workers has to be

[1]For helpful comments and suggestions, I would like to thank Andreas Bernecker, Dirk Engelmann, Botond Kőszegi, Jörg Oechsler, Henrik Orzen, Alexander Paul, Stefan Penczynski, and Sigrid Suetens. I also received helpful comments from participants at seminars in Abu Dhabi, Hamburg, Heidelberg, Mannheim, Munich, the Thurgau Experimental Economics Meeting 2013 (Theem), the Florence Workshop on Experimental and Behavioral Economics 2013, the ESA World Meeting 2013 (Zurich), the 8th Nordic Conference on Behavioral and Experimental Economics 2013 (Stockholm), the Gesellschaft für experimentelle Wirtschaftsforschung 2013 (Helmstedt), the Annual Meeting of the Royal Economic Society 2014 (Manchester).

[2]As outlined by Fehr and Falk (1999, p. 107), examples of these theories are the *implicit contract theory* (see e.g. Baily 1974; Gordon 1974; Azariadis 1975), the *insider-outsider theory* developed by Lindbeck and Snower (1988), *job search models* (see e.g. Mortensen 1986), the *adverse selection model* by Weiss (1980) and the *efficiency wage hypothesis*. For the last approach, some theories focus on a disciplining version of the efficiency wage hypothesis (Gintis 1976; Shapiro and Stiglitz 1984), whereas other theories focus on a fairness version of the efficiency wage hypothesis (Akerlof 1982; Akerlof and Yellen 1990).

adjusted by moderately cutting wages if workers are fair-minded.[3]

How can we reconcile the empirical finding of wage rigidity with the opposite prediction of standard fairness models? In this paper, I will analyze a potential explanation: Workers' fairness considerations might be reference-point dependent. In the literature, there is not only ample evidence that fairness considerations matter in experimental labor markets, but there is also at least some evidence that fairness considerations actually are reference-point dependent (Fehr et al. 2009 survey the literature). An interesting example for this reference dependence of fairness considerations is that Falk et al. (2006) have shown that introducing a minimum wage strongly affects workers' reservations wages (see also Brandts and Charness 2004). For the wage rigidity question, reference dependence of fairness considerations might imply that what workers perceive as a fair wage in recession depends on the wage they previously received out of recession. Already Akerlof (1982) suggested that past wages are one determinant of workers' perception of a fair wage.

Additionally, not related to wage rigidity but supporting the ideas presented so far, Hart and Moore (2008) have claimed that contracts create feelings of entitlement and hence serve as reference points. In their model, Hart and Moore assume that contract parties are driven by self-serving biases and evaluate their final outcome in comparison to their initial contract conditions.[4] If these feelings of entitlement are not fulfilled, that is if one party finally gets less than what it feels entitled to, it will shade, giving rise to inefficiencies. Applying this logic to the wage rigidity setting suggests that contracts concluded initially before recessions might serve as reference points for workers' wage expectations when the economy moves in a recession and these contracts are revised. Workers then evaluate wage cuts against the reference wage of the initial contract conditions and hence perceive wage cuts as unfair, although firms have a good reason to cut wages, the (externally given) recession. In addition to Hart and Moore (2008), also Herweg and Schmidt (2012) argue that contracts serve as reference points, however, not because of self-serving biases but because of loss aversion. Importantly, in this chapter, I will remain agnostic about *why* contracts potentially serve as reference points, whether it is due to loss aversion or due self-serving biases, I just analyze *whether* they actually do in a labor market setting.

The potential importance of reference-dependent fairness considerations has recently also been acknowledged by Eliaz and Spiegler (2013). The authors have argued that implementing wage rigidity in search-matching models (Mortensen and Pissarides 1994) may provide an explanation for the so-called *Shimer puzzle*, namely that standard search-matching models cannot fully account for real-life labor-market fluctuations (Shimer 2005;

[3]Intention-based fairness models (e.g. Charness and Rabin 2002) potentially allow for wage rigidity if one assumes that wage cuts are considered as unkind actions. Importantly, however, these models do not offer any explanation why wage cuts should be considered as unkind. If firms' profits fall due to external circumstances as in a recession, on a first glance, it seems plausible that fair-minded workers should not perceive wage cuts as unkind, because firms have a good reason to cut wages.

[4]Hence, unlike in the setting of Köszegi and Rabin (2006) in which people's expectations determine reference points, in the Hart-and-Moore setting, the status quo determines reference points.

Hall 2005). Importantly, the authors generate wage rigidity in their model by assuming that workers have reference-dependent fairness considerations. They argue that such a modeling approach can explain increased cyclical unemployment volatility even so wages are only rigid for existing workers but not for new hires.[5]

Importantly, however, although data from questionnaire studies (e.g. Kahneman et al. 1986; Bewley 1999) is in line with the hypothesis that the *dependence of fairness considerations on reference points* is one explanation for wage rigidity, rigorous causal evidence is missing so far. This chapter provides such evidence by implementing an experimental labor market in which contracts can serve as reference points, following the idea of Hart and Moore (2008). In this setting, I mainly want to test whether workers really resist wage cuts when their fairness considerations can be reference-point dependent. For this purpose, I also abstract from issues such as unemployment and my design tries to switch off other potential causes of wage rigidity. Thus, the aim of my experiment is not to weigh different possible explanations against each other but to check whether one plausible explanation produces the predicted effect under controlled laboratory conditions.

I implement an experimental labor market based on the gift-exchange game introduced by Fehr et al. (1993). In this game, firms first make a wage offer. Workers then observe these offers and have to decide how much effort to exert. Importantly, however, I deviate from the standard gift-exchange setting by implementing a dynamic component in a similar way to Kocher and Strasser (2011) and Gerhards and Heinz (2012). In my setting, firms have to face recessions (implemented via a negative profit shock) in some periods. In these periods, only the firm's profit is ceteris paribus decreased by a fixed amount.

In this general setting, I implement two main treatments. In my *Baseline* (BASE) treatment, firms and workers are first informed about whether or not a recession (negative profit shock) has occurred and *afterwards* have the ability to conclude a contract. In my *Contracts as Reference Points* (CasRP) treatment, this sequence of events is changed: Every period, workers and firms first conclude a contract *before* they are informed about whether or not a recession has occurred in this period. In case a recession has occurred, the concluded contract can be revised. Otherwise, initial contract conditions determine payoffs. This means that in the CasRP treatment contracts can serve as reference points, whereas this is not the case in the BASE treatment. The question then is whether reference points (triggered by initial contract conditions) in the CasRP treatment lead to workers' resistance against wage cuts and whether we hence observe wage rigidity.

[5] Additionally, wage rigidity has recently played a prominent role in the business cycle literature. It has been argued that the introduction of (real or nominal) wage rigidity into the New Keynesian dynamic stochastic general equilibrium model allows to overcome some well-known shortcomings of the basic model. Riggi (2010) surveys the literature: It has been claimed that incorporating wage rigidity helps to model inflation persistence (Christoffel and Linzert 2010), it helps to better model wage dynamics after a positive technological shock (Zheng and Phaneuf 2007), and it helps to avoid the so-called *divine coincidence*, namely that stabilizing inflation is equivalent to stabilizing the welfare relevant output gap (Blanchard and Gali 2007).

My data suggests that wages are neither completely rigid in the CasRP nor in the BASE treatment. Both in the BASE and the CasRP treatment, wages are cut in recession. Importantly however, wage reductions are higher in the BASE treatment than in the CasRP treatment. Average wage is reduced by 13.1 points (or 21 percent) in the BASE treatment but less than half this amount (6.0 points or 10 percent) in the CasRP treatment. In the CasRP treatment, workers punish wage cuts beyond the mere reduction of the wage level, which provides clear evidence of a reference-point effect. Fairness considerations are reference-dependent and initial contract conditions seem to serve as a reference point for workers' wage expectations.

On the basis of these results, I implement two control treatments. First, the *Wages as Reference Points* (WasRP) treatment tests through which channel contracts serve as reference points, whether an explicit contract conclusion is necessary for the reference-point formation or whether contracts serve as reference points just by determining the out-of-recession wage and conveying this wage to workers. Second, the *Contract as Reference Points treatment with Feedback* (CasRP-F) is very similar to the CasRP treatment. It just slightly improves the information condition of the firm by providing more explicit feedback about how workers react to wage cuts. The question then is whether more information leads to more rigid or even completely rigid wages.

The results for the WasRP treatment show that already knowing the payoff-irrelevant out-of-recession wage in recession is sufficient to create reference points. Surprisingly, an explicit contract conclusion is not necessary. Hence, in my setting, contracts serve as reference points just by determining the out-of-recession wage. A potential explanation for this result is that workers in the WasRP treatment still may have good reasons to perceive the situation *as if* there had been a contract conclusion. Finally, providing more information about workers' behavior to firms (CasRP-F) does not lead to more rigid wages, at least on average.

Although my main result that the reference dependence of fairness considerations provides one explanation for wage rigidity is in line with expectations, it is still surprising to the extent that there are good reasons to believe that a reference-point effect should be much weaker in the laboratory than in the field. In the field, workers get used to their wage over a longer period of time and wage cuts put workers' living standards at risk, which is not the case in the laboratory. Additionally, in the laboratory, it is very clear that moderate wage cuts are a fair action because ceteris paribus only firms' profits decrease in recession. Still, allowing social preferences to be slightly irrational or in other words, allowing for a simple form of context dependence, namely the possibility of reference-dependent preferences, without changing monetary incentives, surprisingly seems to have a huge impact on wages.

An additional contribution of my paper is that it may reconcile conflicting results in the experimental literature about wage rigidity. Fehr and Falk (1999) find evidence in favor

of wage rigidity but only in a static gift-exchange setting with exogenous unemployment. Burda et al. (2005), Kocher and Strasser (2011), and Gerhards and Heinz (2012) all look at a dynamic labor market setting in which firms have an incentive to cut wages in some periods and find results in line with standard (i.e. not reference-dependent) fairness-model predictions. Only Hannan (2005) finds that at least some workers punish wage cuts in a similar setting. Because Hannan's study, however, does not intend to analyze wage rigidity, it is unclear how rigid wages overall are in her study. All these dynamic gift-exchange studies differ in many design features but only Hannan (2005) uses a setting in which a contract is concluded initially before recessions. Hence, only in her setting, contracts might serve as reference points. Importantly, however, she does not provide a control treatment without initially concluded contracts.

In a more general bargaining setting, Gächter and Riedl (2005) test whether entitlements that are not feasible anymore shape negotiation behavior. Although the authors find support for this idea, subjects' fairness judgments in this situation seem to be driven by self-serving biases only to a surprisingly modest degree. Other experimental papers (Fehr et al. 2011; Fehr et al. 2014; Brandts et al. 2012) have tested the precise predictions of the model of Hart and Moore (2008). Importantly, in my experimental design, I only build on the very basic idea of Hart and Moore that contracts create feelings of entitlement and hence serve as reference points. In their model, Hart and Moore do not analyze wage rigidity but why people write long-term contracts. For this purpose, they analyze two different types of contracts, rigid vs. flexible, and their implications for reference points. Because of reference-point formation, a trade-off between contractual rigidity and flexibility arises and Fehr et al. (2011) have experimentally confirmed this trade-off. Bartling and Schmidt (2014) are the first that deviate from this analysis by not comparing different types of contracts but by comparing - as I do - bargaining situations in which initial contracts are renegotiated to bargaining situations in which no ex ante contract was written. Importantly, Bartling and Schmidt do not consider the labor market or wage rigidity but a canonical buyer-seller relationship. In their setting the overall payoff rises in the renegotiation stage and initial contracts constrain sellers (workers) when asking for markups in this stage. In my setting, however, the overall payoff shrinks in the renegotiation stage due to the recession and contracts might constrain firms (buyers) when deciding to pay lower wages. So, it is not clear whether the authors finding that contracts serve as reference points in their setting survives in my setting.[6] Hence overall, the literature has not yet provided a rigorous test of whether reference-dependent fairness

[6] As outlined before, Herweg and Schmidt (2012) additionally provide a theory of contracts as reference points based on loss aversion that is directly applicable to the design of Bartling and Schmidt (2014): Contracts serve as reference points and agents compare gains and losses of renegotiations to this reference point. Because optimally behaving fair-minded agents do not gain in my design and only occur losses, the authors framework is not perfectly applicable to my setting. Hence, my experiment is not designed to discriminate between the explanations of Hart and Moore (2008) and Herweg and Schmidt (2012).

considerations can explain wage rigidity.[7] This chapter provides this test.

The rest of this chapter is organized as follows: Section 3.2 explains the experimental design. Section 3.3 discusses my hypotheses. Section 3.4 provides the experimental results and Section 3.5 concludes.

3.2 Experimental Design and Procedures

In this section, I will first present the experimental design of my two main treatments (Section 3.2.1). Afterwards, I will discuss two control treatments (Section 3.2.2) and the experimental procedures (Section 3.2.3).

3.2.1 Main Treatments: BASE vs. CasRP

My experimental design modifies the traditional gift-exchange setting introduced by Fehr et al. (1993) in at least two important ways. First, in my experiment, subjects play a bilateral gift-exchange game with stranger matching. I implemented a bilateral version of the game because I would like to focus on the behavioral responses of workers to wage cuts, whether workers resist wage cuts or not, deliberately excluding market interactions. Stranger matching is used in order to minimize the importance of reference points other than those induced by initially concluded contracts.[8] The second important deviation from the standard framework is that my gift-exchange setting is a dynamic setting to the extent that firms can be hit by a negative profit shock with a certain probability.

As common in gift-exchange settings, firms make a contract offer by choosing a wage, an integer $w \in \{30, \ldots, 100\}$, and workers who accept a contract offer then have to choose how much effort to exert, $e \in \{1, \ldots, 10\}$. Firms' and workers' decisions about wages and effort levels then determine subjects' monetary payoffs. I implemented a modified version of the payoff functions used by Riedl and Tyran (2005). Firms have the following payoff function:

$$\Pi_F = \begin{cases} 10 \times e - w + 50, & \text{if no profit shock occurs } (p = \frac{2}{3}), \\ 10 \times e - w + 30, & \text{otherwise } (p = \frac{1}{3}). \end{cases}$$

[7] Kube et al. (2013) provide a field experiment in which they find that a wage decrease (without a good reason) damages the work morale. Importantly, however, Chen and Horton (2009) show that providing good reasons for wage cuts can largely eliminate this effect. Hence, it is unclear, whether wage cuts caused by an exogenously given recession really will lead to a decrease in work morale.

[8] A partner matching obviously would have captured reputational concerns and repeated game effects that are present in field labor markets. Besides allowing for various other reference points, however, a laboratory partner matching would potentially have the disadvantage of highlighting the role of reputational concerns too strongly: Due to the short length of the experiment, the importance of future periods is very salient in the lab, whereas in the field, workers resistance against wage rigidity might be driven by emotions and the importance of future periods might be less salient due to the longer time horizon.

Workers' payoffs are given by

$$\Pi_W = w - c(e).$$

Here, F denotes the firm, W denotes the worker, and $c(e)$ is the cost of effort that is an increasing function of e. In my experiment, I used a standard cost-of-effort function shown in Table 3.1. Workers who reject their contract offer get a payoff of $\Pi_{W_0} = 20$ whereas their assigned firms get a zero payoff, $\Pi_{F_0} = 0$. Additionally, firms also earn zero profits if their contract offers have been accepted but effort levels had resulted in a negative payoff for the firm.

Table 3.1: Effort Levels and Cost of Effort

e	1	2	3	4	5	6	7	8	9	10
c(e)	0	1	2	4	6	8	10	12	15	18

The crucial feature of my design is that the negative profit shock mimicking a recession only has an effect on firms' profits. It ceteris paribus reduces a firm's profit by 20 points. Importantly, each period, one third of all firms are hit by a negative profit shock. So, unlike in the field, only some firms have to face recessions in each period and not all firms in some periods. Because each worker-firm match, however, is not informed about the results reached by other matches, the design reflects a situation in which we have various independent islands, each with a worker-firm match, and on one third of these islands the firm is hit by a recession. I implemented the recession as described mainly because of a potential follow-up paper for which such a recession implementation would be very helpful. Because of the bilateral structure of the game, however, these two ways of implementing a recession should not lead to qualitatively different results.[9]

Why did I implement these particular parameters? Importantly, empirical evidence by questionnaire studies - as outlined in the Introduction - suggests that wages overall are at least too rigid to avoid an increase in unemployment. However, these studies (see e.g. Bewley 1999, or as additional evidence Kahneman et al. 1986) also find that some firms are able to cut wages substantially. Wage cuts seem to be possible if a firm can credibly claim that its bad economic situation is caused by the recession and if its economic situation is so bad that the firm's existence is at risk. Workers seem to be willing to make major concessions regarding their wage if they fear that they might loose their job by not making these concessions.

Regarding my experimental design, these considerations suggest that contracts might not serve as reference points under all possible payoff structures or parameters. The

[9] Additionally, my way of implementation increases the variability of when firms and workers are hit by a recession, and it captures to some extent the workers' view on recessions. For example Bewley (1999) finds that workers mainly focus on their own company and know only very little about other companies' economic situations in recession.

question is whether there are reasonable parameters leading to feelings of entitlement and hence wage rigidity. Trial sessions were run in order to identify conditions under which a reference dependence of fairness considerations occurs. The main guiding principle for my parameters derived from these trial sessions is that firms should experience a decrease in profit during a recession but that these reductions should not be so large that firms actually experience negative payoffs. As seen in the discussion about the field, such losses provide a strong argument in favor of wage cuts that are potentially not questioned by reference points.[10] In the following, I will shortly discuss how some aspects of my payoff functions create an environment in which it might (if at all) be reasonable to expect that reference points have an impact on wages:

- Because the gift-exchange relationship favors workers (especially with strangers matching), I implemented some extra money (50 points) in the firm's payoff function in order to strengthen the firm and to avoid low payoffs of firms. Additionally, I implemented that firms even only earn zero profits if the worker's effort level had resulted in negative payoffs for the firm.

- For implementing a recession through a negative profit shock, there are at least two possible alternatives: Either the productivity parameter with which the effort level is multiplied could be decreased or the firm's extra money could be reduced by a fixed amount. Importantly, cutting the productivity parameter makes either high wages unsustainable if the parameter is decreased by a large amount or it does not have a real impact on small wages if the parameter is decreased only by a small amount.[11] One way of avoiding this problem would be to allow for different reductions of the productivity parameter for different wage levels. For the ease of a simpler environment, however, I choose to reduce the extra money in recession by a fixed amount. Hence both high and low wages are reduced by the same amount.[12]

- I implemented a fairly high minimum wage of 30 in order to have a low risk-free payoff of 10 points that firms can ensure themselves in recession. A lower minimum wage would have led to a higher risk-free payoff in recession, increasing the likelihood of wage cuts. Additionally, workers get an unemployment payment ($\Pi_{W_0} = 20$) that is lower than the minimum wage. This provides workers with a strong incentive to

[10] Hart and Moore (2008) argue that self-serving biases create feelings of entitlement and that hence contracts serve as reference points. In a situation, in which firms face losses, such kinds of self-serving biases might not be present any more.

[11] Kocher and Strasser (2011) cut their productivity parameter from 10 to 5 and hence observe unsurprisingly wage cuts in all their treatments and in their treatment with stranger matching firms frequently make losses, especially in the initial rounds of the experiment.

[12] As a profit shock, I choose an amount of 20 because of two considerations: First, the profit shock has to be big enough to create statistically distinguishable differences between the treatments. Second, it has to be small enough so that in principle not cutting wages can be a profitable strategy for firms if workers resist wage cuts.

accept every contract offer, and it also helps to avoid a different frequency of workers between treatments that reject contract offers. Hence, in my design, accepting or rejecting an offer is not so much about evaluating whether a contract offer is good or bad, but it is more about intensifying feelings of entitlement in the CasRP treatment through an explicit contract conclusion. Finally, the minimum wage is not adjusted in recession to avoid a potential experimenter demand effect.

In sum, my payoff functions try to create a setting that gives fairness considerations a reasonable chance to be reference-dependent but that still realistically reflects field conditions. Importantly, payoff functions and parameter are the same for all treatments, making wage cuts (due to monetary incentives) a priori as likely in all treatments.

The major difference between the CasRP and the BASE treatment is that in the first treatment contracts concluded before recessions potentially serve as reference points for workers' wage expectations in recession, whereas such initially concluded contracts do not exist in the BASE treatment. Additionally, I think that the CasRP treatment captures an important aspect of the primary labor market in a stylized way: In the primary labor market, long-term relationships are common, and hence previous out-of-recession wages may influence wages in recession. On the contrary, the BASE treatment might reflect in a stylized way that in the secondary labor market short-term relationships between firms and workers are common, and hence previous wages do not have a large effect on current wages.

The details of the experimental design are the following: It is common to both treatments that subjects play in matching groups of 12 people with 6 firms and 6 workers.[13] In each period, one firm and one worker are randomly matched and subjects play 18 periods.

BASE treatment

In the BASE treatment, each period has the following time structure:

- *Stage 1*: At the beginning of the first stage, two out of six firms are randomly selected. These two firms are hit by a negative profit shock. Knowing whether they have been hit by such a shock or not, firms then make a contract offer to workers by choosing a wage, $w \in \{30, \ldots, 100\}$.

- *Stage 2:* This stage depends on the previous one. In case the assigned firm is not hit by a negative profit shock, workers first have to decide whether they want to accept their contract offer or not. If they accept, workers have to choose their effort level, $e \in \{1, \ldots, 10\}$. In case the assigned firm is actually hit by a negative profit

[13] In some sessions due to no shows, subjects played in matching groups of 10 people with 5 firms and 5 workers.

shock, workers only have to choose their effort but cannot reject the contract offer. This difference is implemented in order to ensure comparability with the CasRP treatment.

- *Stage 3:* Firms and workers get to know their profits, and firms are also informed about their workers' effort if their contract offer has not been rejected.

CasRP treatment

The CasRP treatment implements the possibility that contracts concluded before recessions serve as reference points. Hence, the sequence of events and especially the time at which subjects are informed about whether or not a negative profit shock has occurred is changed. In the CasRP treatment, each period has the following time structure:

- *Stage 1:* Firms make a contract offer to their workers by choosing a wage, $w_1 \in \{30, \ldots, 100\}$. Afterwards, these contract offers are communicated to workers who have to decide whether they want to accept them or not. If they accept, workers have to choose their effort level, $e_1 \in \{1, \ldots, 10\}$.

- *Stage 2:* At the beginning of the second stage, two out of six firms are randomly selected. These two firms are hit by a negative profit shock and only these two firms are allowed to change the initial contract conditions by choosing a new wage, $w_2 \in \{30, \ldots, 100\}$. Because an accepted contract is binding for workers, they cannot reject these wage changes. Nonetheless, affected workers are allowed to adjust their initial effort by choosing a new effort level, $e_2 \in \{1, \ldots, 10\}$, implementing a punishment device for workers. Firms (and assigned workers) who are not hit by a profit shock cannot change the initial contract conditions and hence do not have to make any decisions in this stage.

- *Stage 3:* Firms and workers get to know their profits, and firms are also informed about their workers' *relevant* effort if their contract offer has not been rejected. This means that firms only learn about e_1 in case a negative shock does not occur, and in case such a shock occurs, firms only learn about e_2.

Payoff functions and parameters are the same as in the BASE treatment. In case that no profit shock occurs, wages and effort levels of stage 1, w_1 and e_1, determine payoffs. Otherwise, wages and effort levels of stage 2, w_2 and e_2, determine payoffs. Hence, in the CasRP treatment, firms make a contract offer for which both firms and workers know that the contract conditions can be changed by the firm when a recession occurs.

My two main treatments provide a test whether the reference dependence of fairness considerations provides one explanation for wage rigidity. Reference points are implemented following the idea of Hart and Moore (2008) that contracts serve as reference points.

Importantly, subjects in the CasRP treatment not only know the wage they would have gotten out of recession also in recession (which subjects do not know in the BASE treatment), but subjects also explicitly agree to this wage (or the initially concluded contract). I deliberately chose to implement the workers' explicit agreement to the initial wage and hence change two things at the same time between treatments because I conjectured that an explicit agreement of both contract parties to the contract would be necessary for contracts serving as reference points. The explicit agreement to the wage offer potentially provides workers with an additional excuse to punish wage cuts and reinforces workers' self-serving biases: Workers who explicitly agreed upon a contract might argue for themselves that wage cuts are not legitimate because they violate the initial agreement with their firm. Although this argument seems objectively insufficient to the extent that workers are aware that firms have the possibility to change the initially concluded contract if a recession occurs and more importantly that there is still an objective reason to cut wages (the recession), it might convince at least some workers to punish wage cuts.[14] My first control treatment tests whether an explicit agreement is really necessary for observing workers' resistance against wage cuts.

In a recession, the contract revision is implemented in a very extreme way, following Fehr et al. (2014). Firms can revise their worker's wage as they like and workers cannot reject the new wage but are able to change their effort level. Fehr et al. (2014, p. 3) have argued that this implementation provides "a powerful stress test" for contractual reference points because "the easier it is to change a contract the less likely it is to serve as a reference point." On the other hand, one may argue that the fact that firms have all the power in changing an established contract may bias my results in favor of workers' resistance to wage cuts because workers may perceive the procedure as unfair. In my first control treatment, however, such an established contract does not exist and I control for this potential problem.[15]

Finally, some short remarks about what the term *wage cut* can mean in the two different treatments? In the BASE treatment, we basically only observe an *indirect* wage cut. We can compare average wages of firms in recession with average wages of firms out of recession. But by construction of the treatment, we do not observe how individual firms

[14] Another feature that potentially intensifies the explicit contract conclusion aspect is that in the CasRP treatment subjects also choose an out-of-recession effort level even if a recession occurs, which is not the case in the BASE treatment. Already choosing such an effort level might additionally foster feelings of entitlement regarding the initial contract conditions.

[15] Additionally, implementing a workers' acceptance decision in the revision stage would have led to other problems: In the case of implementing the workers' rejection decision such that it leads to no employment relationship between workers and firms, rejecting high wages that have been cut substantially but not to the minimum wage, might not be a very reasonable thing to do because of the very high costs associated for workers. Workers would finally then only earn $\Pi_{W_0} = 20$. In the case of implementing the workers' rejection decision such that initial contract conditions then determine payoffs, the BASE treatment could not provide a meaningful comparison treatment because workers' decision problem in recession would then be different in this treatment.

cut wages in one period because firms make their wage decision with the knowledge about whether a recession has occurred or not. In the CasRP treatment, however, we can not only observe this kind of *indirect* wage cut measure, but we can also observe a *direct* wage cut by comparing wages from stage 1 and stage 2 in case a recession occurs. Of course, there is no theoretical reason why these *direct* wage cuts should be different from *indirect* wage cuts. Nonetheless, smaller difference could randomly occur, and hence I will use both measures in the results section.

3.2.2 Control Treatments: WasRP and CasRP-F

I implemented two additional control treatments. The first control treatment is the *Wages as Reference Points* (WasRP) treatment. This treatment tries to disentangle which of the two main aspects changed from the BASE to the CasRP treatment potentially leads to a reference dependence of fairness considerations. Hence, it tests through which channel contracts serve as reference points. Is an explicit contract conclusion necessary to create workers' resistance against wage cuts or do contracts just serve as reference points by determining the out-of-recession wage and conveying this information to workers? Concerning the sequence of events, this treatment is a combination of the BASE and the CasRP treatment:

- *Stage 1*: Firms choose a contract offer (by specifying a wage) for the case that no negative profit shock occurs. Then, the computer selects two out of six firms. These firms are hit by a negative profit shock and are allowed to adjust their contract offer for the recession.

- *Stage 2:* If the assigned firm is not hit by a negative profit shock, workers first have to decide whether they want to accept their (out-of-recession) contract offer or not. If they accept, they have to choose their effort. If the assigned firm is hit by a negative profit shock, workers only have to choose their effort but cannot reject their (recession) contract offer. Importantly, in the second case, workers are also informed about the out-of-recession contract offer (wage) that is irrelevant for payoffs.

- *Stage 3:* Firms and workers get to know their profits, and the firms are also informed about their workers' *relevant* effort if their contract offer has not been rejected.

Importantly, the WasRP treatment is fairly artificial. In this treatment, workers basically know the (payoff-irrelevant) out-of-recession wage also in recession although they have not explicitly concluded a contract with their firm before the recession. It still seems reasonable that workers may still perceive their situation *as if* a contract had been concluded between them and their firms[16], but the question remains whether this

[16]In each period, each firm is matched with one worker. Hence, firms first decide about a out-of-recession wage offer for their individually matched workers, and those firms facing recession afterwards also make

perception is enough to invoke a sufficient level of workers' self-serving biases leading to wage rigidity or whether for this purpose an explicit agreement of workers to the contract is still necessary. But even if the knowledge of the out-of-recession wage is already enough to generate wage rigidity, contracts might still serve as reference points, just not by an explicit contract conclusion but because they determine the worker's personal out-of-recession wage and convey this information to workers. In the field, workers will in most cases only know their personal out-of-recession wage if they have concluded a contract.

My second control treatment is the *Contracts as Reference Points treatment with Feedback* (CasRP-F). This treatment is very similar to the CasRP treatment. Only the information condition of the firm is slightly changed. Hence, only *Stage 3* of the game is different from the CasRP treatment:

- *Stage 3:* Firms and workers get to know their profits, and firms are also informed about their workers' effort if their contract offer has not been rejected. This means that firms only learn about e_1 in case a negative shock does not occur, but in case such a shock occurs, firms learn about both e_1 and e_2.

The reason to implement this control treatment is the following: To ensure comparability between the BASE and the CasRP treatment, firms are only informed about their workers' relevant effort levels in the CasRP treatment. So, if a recession occurs, firms are only informed about their workers' recession effort level, not about their workers' effort level out of recession. For this reason, it might not be straightforward for firms to infer how workers react to wage cuts. A potential problem is that workers may punish wage cuts but that firms are unable to detect this punishment either because this punishment is not very strong or because they cannot distinguish punishment from egoistic behavior. Hence, if workers punish wage cuts, it is possible that the difference between the BASE and the CasRP treatment underestimates the real difference caused by reference points. Especially, if a treatment difference between the BASE and the CasRP treatment is found but wages are not completely rigid in the CasRP treatment, we may wonder whether more information about workers' behavior, as provided in the CasRP-F treatment, leads to complete wage rigidity.[17] In the CasRP-F treatment, firms can directly infer how workers adjust their effort level in case firms cut wages (or increase wages).[18]

a recession wage offer. Workers in recession may then still feel entitled to their payoff-irrelevant out-of-recession wage because this wage is not an arbitrary wage that has nothing to do with them, but it is their personal out-of-recession wage they would have gotten if a recession had not randomly occurred.

[17] Due to the construction of the BASE treatment, such kind of control treatment cannot be implemented for the BASE treatment.

[18] Another possibility is that we observe wage rigidity in the CasRP treatment, although workers do not punish wage cuts. This might happen if firms make wrong inferences about workers' behavior due to the bad information conditions and wrongly assume that workers punish wage cuts, although this is not the case. In this case, wages might actually be more flexible in the CasRP-F treatment.

3.2.3 Subjects, Payments and Procedures

Sessions lasted on average between 75 and 90 minutes and took place at the mLab at the University of Mannheim in Spring and Autumn 2013. Overall, 278 subjects participated in the experiment and earned on average 20 EUR. For all four treatments, six sessions were run. For the CasRP treatment, five sessions included 12 participants, and one session included 10 participant. For the BASE and the CasRP-F treatment, four sessions included 12 participants, and two sessions included 10 participant. For the WasRP treatment, all sessions included 12 participants.

In order to make subjects familiar with the computer screens, three additional trial rounds were conducted. In these rounds, firms could not be hit by a negative profit shock (recession). Recruitment was done by ORSEE, Greiner (2004), and the experiment was programed in z-Tree (Fischbacher 2007). Experimental instructions for all treatments can be found in Appendix B.4.

3.3 Hypotheses

In this section, I will discuss three different hypotheses regarding behavior in the experiment.[19] In the text, I will only provide informal arguments for these hypotheses. Appendix B.2 provides more formal arguments based on preferences of inequity aversion (Fehr and Schmidt 1999) and the framework of Hart and Moore (2008).

Assuming common knowledge of perfectly rational and selfish agents, the game theoretic solution for the gift-exchange game is straightforward. Workers do not have an incentive to provide more than minimal effort, $e = 1$. Firms anticipate this and choose the minimum wage for workers, $w = 30$. In recession, firms' and workers' behavior does not change, although a firm's profit is reduced when the firm is hit by a recession. Firms cannot cut wages in recession because they already pay the minimum wage out of recession. This outcome is the same for all treatments because material incentives do not change between treatments. Obviously, it would have been possible to make wage cuts an equilibrium prediction by lowering the minimum wage in recession.[20] However, as outlined before, in order to avoid any kind of experimenter demand effect that lowering wages is a reasonable thing to do, I did not implement a reduction of the minimum wage in recession. Overall, the gift-exchange literature has shown that the subgame perfect equilibrium, assuming selfish agents, is not a good predictor for this game. Hence, whether wage cuts happen in such an equilibrium is not really an empirically interesting question, and thus my design was not tailored to allow for such equilibrium wage cuts.

[19] Hypothesis 1 and 2 are actually fairly similar to the Hypothesis 2 and 3 of Bartling and Schmidt (2014), only that my hypotheses capture my labor market setting.

[20] Such a lower minimum wage might reflect that in the field in a recession high unemployment rates potentially allow firms to hire new entrants at lower wages.

Fairness models have shown to predict participants behavior in the gift-exchange game more accurately. These models assume that at least some people are not only motivated by material self-interest but also by the material payoffs of others. For these people, fairness models, such as inequity aversion by Fehr and Schmidt (1999) or social welfare preferences by Charness and Rabin (2002), then assume that others people's welfare enters their utility function. Regarding my first hypothesis, I will focus on so-called outcome-based fairness models because they allow for a clear-cut prediction.[21] For these models, a recession makes an adjustment of the split of surplus between a firm and a worker necessary because only the firm is hit by a negative profit shock, whereas the worker's payoff ceteris paribus remains the same. Assuming that workers are sufficiently fair-minded (whereas firms may or may not have standard preferences)[22], the following hypothesis holds:

> ***Hypothesis 1*** *[Outcome-based Social Preferences]*: In both main treatments, BASE and CasRP, firms moderately adjust their wages downward if they are hit by a negative profit shock. This adjustment is not different between the two treatments. Controlling for wages, workers exert more effort in recession than out of recession. Again, there is no difference between treatments.

Assuming that workers do not only care about other people's welfare in their utility function but that contracts also serve as reference points, as suggested by Hart and Moore (2008), changes the prediction. Contracts can serve as reference points to the extent that initial contract conditions create feelings of entitlement. Here, as outlined before, I remain agnostic *why* contracts actually serve as reference points, whether self-serving bias are important or whether loss-aversion leads to this effect. In CasRP treatment, if a firm is hit by a negative profit shock and cuts wages, the worker feels entitled to the wage of the initially concluded contract and shades by lowering effort.[23] Firms should anticipate this and not lower wages. In the BASE treatment, however, contracts are not concluded before recessions and hence firms should freely adjust wages downward. Although we might not expect such clear cut result, we should at least expect the following:

[21] Intention-based fairness models may predict wage rigidity if wage cuts are assumed to be unkind, but crucially, these models do not offer an explanation why wage cuts are perceived as unkind despite the fact that the random profit shock is externally given. Both types of fairness models predict for the gift-exchange game that wages and effort should be above the minimal level if the fraction of those agents who also care about other agents' payoff is not too small. Additionally, there should be a positive relation between wage and effort.

[22] Assuming additionally that also firms are fair-minded complicates the analysis but does change the results qualitatively.

[23] Following Hart and Moore (2008), I assume that workers feel basically fully entitled to the status-quo wage. Assuming, however, that workers feel entitled to their rational expectations, following Köszegi and Rabin (2006), would complicate the analysis quite a lot but not qualitatively change predictions because even when expectations serve as reference points wage cuts are still perceived as reference point violations. With expectations determining the reference point, workers' reference points would incorporate both the recession wage and the out-of-recession wage with the appropriate probability. Importantly, if the recession wage is adjusted downward, workers still compare this adjusted wage partially with the out-of-recession wage. Hence, any wage cut would still be a reference-point violation.

> ***Hypothesis 2*** *[Contracts as Reference Points]*: In the CasRP treatment, wage cuts are not as large as in the BASE treatment. Hence, wages are more rigid in the CasRP than in BASE treatment. Controlling for wages, workers exert more effort in recession than out of recession in the BASE treatment. For the CasRP treatment, a reference-point effect is observed. Controlling for the wage level, effort is lower for wage cuts than for stable (or increased) wages.

If firms cannot adjust wages in the CasRP treatment as much downward as they do in the BASE treatment, these firms may already want to reduce the out-of-recession wage in the CasRP treatment in order to be able to pay lower wages also in recession. The so-called *implicit contract* literature (see Baily 1974; Gordon 1974; Azariadis 1975) has argued that due to risk-aversion workers prefer stable wages over fluctuating wages. Hence, an implicit contract may arise in which firms do not cut wages but adjust the long-run (out-of-recession) wage downward in order to compensate for not adjusting wages downward in recession. In my setting, workers resistance against wage cuts in the CasRP treatment may hence work as an enforcement device for an implicit contract between workers and firms, as suggested by Bull (1987) and Newbery and Stiglitz (1987). So, if we observe a wage reduction in the CasRP treatment out of recession compared to the BASE treatment, this has some kind of *implicit-contract* flavor. In principle, the recession's probability of 1/3 is low enough to make such a strategy suboptimal in my setting if workers are fair-minded but firms still have standard preferences. Because not cutting wages, however, makes workers in recession better off than firms, inequity-averse firms that want to avoid this inequality may adjust wages already out of recession, as suggested by the implicit contract literature. I do not offer a prediction here but ask the following question:

> ***Open Question 1****:* Do firms in the CasRP treatment adjust the out-of-recession wage downward in order to be able to pay lower wages in recession?

Additionally, I am interested through which channel contracts potentially lead to feelings of entitlement that result in more rigid wages. Is an explicit contract conclusion necessary to create workers' resistance against wage cuts, or do contracts serve as reference points by determining the out-of-recession wage and conveying it to workers? Is it sufficient that workers potentially perceive their situation as if a contract had been concluded between them and their firms, or do we need an explicit agreement between workers and firms? On the one hand, one may conjecture that only an explicit agreement of workers is sufficient to invoke their self-serving biases: Only because they explicitly agreed with a contract offer, they might argue for themselves that a violation of this agreement has to be punished, although the recession provides a good reason to moderately cut wages. On the other hand, if the behavioral forces underlying the reference dependence of fairness considerations are

fairly strong, workers' mere perception of the situation as if a contract had been concluded might be sufficient to lead to wage rigidity. I consider the channel as an open question:

> ***Open Question 2*** *[Wages as Reference Points]*: Does already the knowledge of the payoff-irrelevant out-of-recession wage create feelings of entitlement, or is an explicit contract conclusion a necessary condition for resistance against wage cuts?

If wages in the WasRP treatment are as rigid as in the CasRP treatment, the (payoff-irrelevant) out-of-recession wage seems to be enough to create feelings of entitlement. If on the contrary, wages in the WasRP treatment are as flexible as in the BASE treatment, an explicit contract conclusion seems to be necessary to create such feelings of entitlement.

Finally, regarding the CasRP-F treatment, one might conjecture that an information problem might exist in the CasRP treatment, as outlined before. Even if workers punish wage cuts, firms may still have difficulties of detecting weak forms of punishment because of the low information environment and the stranger matching. Hence, if an information problem exists and if firms are provided with the reaction of their workers to their wage adjustments (as it is done in the control treatment CasRP-F), we should expect lower wage cuts compared to the CasRP treatment if we assume that workers indeed punish wage cuts substantially.

> ***Hypothesis 3*** *[Contracts as Reference Points with Feedback]*: Due to the better information conditions, wage cuts in the CasRP-F treatment are less pronounced than in the CasRP treatment.[24]

An additional aspect is that not only firms may change their behavior in the CasRP-F treatment, but also workers may alter their behavior because they may recognize that they have a fairly good punishment device for firms that cut wages in this treatment, potentially leading to more heavy punishment.

3.4 Results

In this section, I present my experimental results. I first provide an overview of the results for the two main treatments, BASE vs. CasRP (Section 3.4.1). Afterwards, I analyze why a treatment difference between the BASE and the CasRP treatment is observed (Section 3.4.2). Finally, I present the results for the control treatments (Section 3.4.3).

[24] If workers on the contrary do not punish wage cuts (or only to very limited degree) and if we still observe wage rigidity in the CasRP treatment, this might also be due to an information problem. Firms might overestimate workers' punishment. Then, of course, more information should lead to more flexible wages.

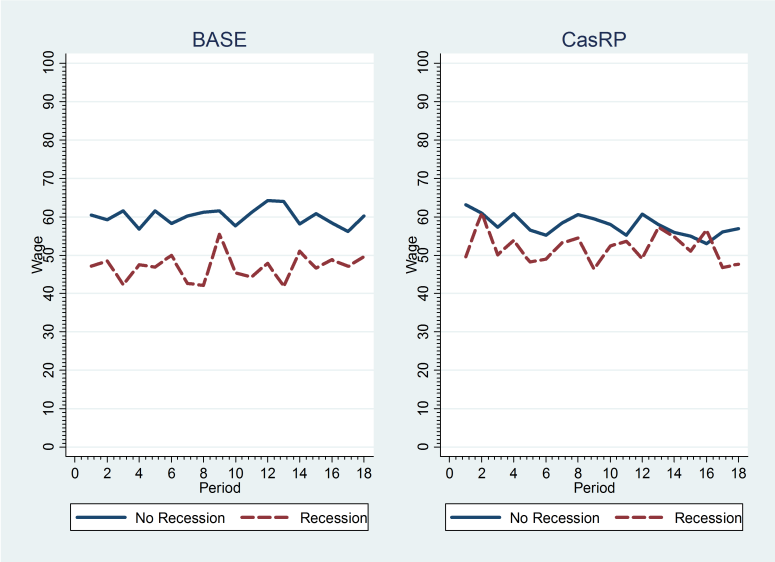


Figure 3.1: Average Wages - BASE vs. CasRP

3.4.1 BASE vs. CasRP: Overview and Firms' Behavior

Figure 3.1 presents average wages over time, separated for firms that face a recession (negative profit shock) and those firms that do not face a recession. On the left, wages for the BASE treatment and on the right wages for CasRP treatment are shown. Overall, it seems that in both treatments, at least on average, wages are lower in recession than out of recession. Additionally, wage cuts seem on average to be smaller in CasRP treatment than in the BASE treatment, although this difference is not overwhelmingly large. Importantly, however, one has to bear in mind that the negative profit shock in recession is only 20 points. If workers and firms split this shock equally (as they should do if workers are fair-minded), we should expect that wages only decline by 10 points. Hence, even if we observed completely rigid wages in the CasRP treatment, the treatment difference would only be 10 points. Compared to this benchmark, the difference between treatments does not seem to be small.

As a first overview, Figure 3.2 also presents average effort, following the same structure as Figure 3.1. At least on average, effort in recession is higher in CasRP treatment than in the BASE treatment, although this difference is not overwhelmingly large. Additionally, this difference might of course just be a mere reflection of the fact that on average wages are not cut as much in the CasRP treatment as in the BASE treatment. I will take a more detailed look at effort, using appropriate regression analysis, in the next section.

In the remainder of this section, however, I will first analyze firms' behavior. For this purpose, Table 3.2 provides average wages in and out of recession for my two main treatments.[25] Additionally, the table provides two measures of the wage difference in and

[25]In all the analysis so far and still to come, I will focus on wages and effort levels that emerge when workers have accepted a wage offer. First, when a wage offer is rejected, the corresponding effort level

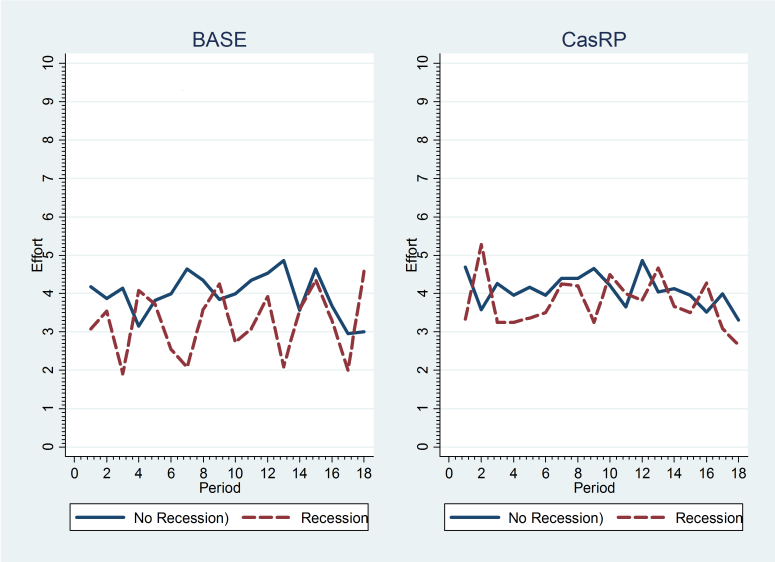


Figure 3.2: Average Effort - BASE vs. CasRP

out of recession. First, it provides the raw difference between average wages out of and in recession, just called *wage cut*. Second, it provides the *relative wage cut*. This variable just measures the *wage cut* relative to the wage paid out of recession (on an individual level).[26] In principle, the (raw) *wage cut* seems to be the suitable wage difference measure: All firms are hit by a negative profit shock of 20 points in recession and hence might want to adjust their wage downward e.g. by 10 points, independent of whether a firm has paid a high or low wage out of recession. Obviously, however, firms paying high wages have much more scope to adjust their wages downward compared to firms paying low wages. This is especially true because the minimum wage of 30 restricts downward wage adjustment, especially for small wages. For this purpose, I provide the *relative wage cut* measure as a complementary measure.

The main result that Table 3.2 shows is that there is a treatment difference in line with the CasRP hypothesis: Average wages are cut in recession by 13.1 points in the BASE treatment whereas averages wages are only cut by 6 points in in the CasRP treatment. Because differences in average wages out of recession are fairly small across treatments,

is missing. Second, *direct* wage cuts only make sense if wage offers have been accepted otherwise the recession wage would be missing. Most contract offers are accepted because not accepting an offer leads to a loss of at least 10 points due to the high minimum wage. Acceptance rates range from 95 percent (BASE) to 99 percent (WasRP), with CasRP and CasRP-F inbetween with 98 percent. Including, however, not accepted wage offers in the analysis of wages leads to very similar results, no systematic differences emerge: All (at least marginally) significant results of non-parametric tests remain at least marginally significant.

[26] More precisely, there are two ways of measuring the wage cut for the CasRP treatment in Table 3.2, as outlined in Section 3.2.1. Because the table provides (payoff-relevant) average wages in and out of recession, it provides the *indirect* wage cut and the *indirect* relative wage cut for the CasRP treatment. The *direct* wage cut and the *direct* relative wage cut are, however, even slightly smaller. The direct no-recession wage in the CasRP treatment is 57.0. Hence, the direct wage cut is 5.1 and the direct relative wage cut is 0.09.

Table 3.2: Average Wages: BASE vs. CasRP

	Wage		Wage Cut	Relative
	No-Recession	Recession		Wage Cut
BASE	60.1	47.0	**13.1**	0.21
CasRP	57.9	51.9	**6.0**	0.10

this difference at least on average translates in a comparable difference in the relative wage cut measure. Importantly, fluctuations in average wages across sessions are fairly high.[27] One reason why this is the case is that matching groups include only 12 people (or 10 people with no-shows). With these small numbers, a group's composition, e.g. how many free-riders (egoistic workers) are present, might have a huge influence on how generous wages are paid. For this reason, I am not able to detect a statistically significant difference in average wages between treatments out of recession, even when testing with more power on the individual level (One-sided Wilcoxon rank sum test - *Direct* no-recession wage: $p = 0.262$, *Indirect* no-recession wage: $p = 0.380$).[28] Also the difference between wages in recession between treatments is only marginally significant when testing on the individual level (One-sided Wilcoxon rank sum test: $p = 0.075$).[29] Hence, I will in the following focus on whether the *wage cut* and the *relative wage cut* are statistically different between treatments according to non-parametric tests. Note, however, that although statistically not significant according to non-parametric tests, wages out of recession are on average lower in the CasRP treatment. A finding that will be statistically confirmed in a regression analysis presented below.

The first question, I would like to answer is whether wages in and out of recession are different within treatment. To analyze this question with non-parametric tests, I will use both *direct* and *indirect* wage cut measures for the CasRP treatment, and all non-parametric tests will be one-sided because both hypothesis 1 and hypothesis 2 provide a directed prediction. Wilcoxon signed rank tests confirm the results of Table 3.2: The difference in the BASE treatment is significant both for testing on the session and on the individual level (Session: $p = 0.014$; Individual: $p = 0.000$). The same is true for the

[27]For example, I observe the following session means of average wages out of recession for the BASE treatment: 48.2, 53.4, 63.0, 63.2, 63.6, 68.9. For the CasRP treatment, I observe: 48.6, 50.0, 56.2, 58.0, 64.9, 68.9.

[28]When analyzing this data with non-parametric tests, one has to decide whether to use individuals or sessions as independent observations. Although my observations are really only independent at the session level (each session as one observation), I will also report non-parametric tests at the individual level (each subject as on observation) for completeness. Compared to session level data, the sample size of course increases quite lot when using individual data, leading to a higher power of the non-parametric test. On the other hand, results might be biased because observations are not really independent on the individual level, as outlined before.

[29]Tests on the session level do not confirm a difference for wages out of recession (one-sided Wilcoxon rank sum test - *Direct* no-recession wage: $p = 0.315$, *Indirect* no-recession wage: $p = 0.374$) and they even do not confirm a difference in recession (one-sided Wilcoxon rank sum test: $p = 0.186$).

difference in the CasRP treatment for both methods of measuring the wage cut (*Direct* wage cut - Session: $p = 0.014$; Individual: $p = 0.001$. *Indirect* wage cut - Session: $p = 0.023$; Individual: $p = 0.000$). This gives me my first result:

> ***Result 1:*** In both treatments, wages are cut in recession. This means that wages are neither in the BASE treatment nor in the CasRP treatment completely rigid.

Hence, the question remaining is whether wages are at least more rigid in the CasRP treatment than in the BASE treatment, as the CasRP hypothesis suggests. One-sided Wilcoxon rank sum tests confirm this hypothesis. Using *direct* wage cuts for the CasRP treatment, the (raw) *wage cut* is significantly different both for testing on the session and on the individual level (Session: $p = 0.013$; Individual $p = 0.005$). The same is true for the *relative wage cut* (Session: $p = 0.013$; Individual: $p = 0.003$). Using *indirect* wage cuts as a control, leads to similar results: On the session level and on the individual level, both measures remain significant at the conventional level (*Wage cut* - Session: $p = 0.019$; Individual: $p = 0.025$. *Relative wage cut* - Session: $p = 0.013$; Individual: $p = 0.020$). Overall, non-parametric tests provide support for the CasRP hypothesis. Wages seem to be more rigid in the CasRP treatment than in the BASE treatment.

As shown before, non-parametric tests, however, do not provide support for the claim that firms already adjust their wages downward in the CasRP treatment out of recession. To further analyze this point and to provide additional support for the other results presented so far, Table 3.3 provides panel random effects (RE) and fixed effects (FE) regressions with wages as the dependent variable and different specifications. These regressions do not just use session or individual averages as the non-parametric tests presented so far but make use of all available data.[30] In all regressions, I control for time trends, standard errors are given in brackets, and clustering is done at the session level, as the level of independent observations. Regressions (1)-(3) are random effects regressions in which a time-invariant treatment dummy (*CasRP-dummy*) is included, whereas regression (4) uses fixed effects as a control.

In regression (1), the wage is only explained by a constant, a dummy for the CasRP treatment, and a recession dummy. This recession dummy is highly significant, telling us that on average over both treatments wages are significantly lower in recession than out of recession. Regression (2) adds an interaction dummy between the treatment dummy and the recession dummy. It tells us that in the BASE treatment wages are on average cut by 11.5 points whereas in the CasRP treatment this wage cut (in recession) is 5.8 points smaller. This difference is significant at 5%-level and hence we have further evidence that

[30]For the CasRP treatment, in recession, I not only make use of stage 2 wages but also utilize stage 1 wages in the regression analysis. This is done in order to make use of all data available. Additionally, it seems appropriate to the extent that stage 1 wages are fully incentivized because firms do of course not know whether a recession will occur or not when making the stage 1 decision.

Table 3.3: Panel Regressions on Wages, BASE & CasRP

	(1) RE	(2) RE	(3) RE	(4) FE
CasRP-dummy	0.110	-1.649	-3.241**	
	(4.137)	(4.229)	(1.392)	
Recession-dummy	-8.365***	-11.49***	-10.73***	-10.85***
	(1.437)	(1.589)	(2.367)	(2.181)
Rec x CasRP		5.810**	6.161**	5.560**
		(2.305)	(2.521)	(2.402)
Wage_{t-1} (No-Rec)			0.435***	0.163***
			(0.0610)	(0.0359)
Wage_{t-1} (Rec)			0.483***	0.176***
			(0.0653)	(0.0386)
Effort_{t-1} (No-Rec)			3.052***	2.474***
			(0.359)	(0.282)
Effort_{t-1} (Rec)			3.180***	2.648***
			(0.436)	(0.322)
First wage			0.0573	
			(0.0386)	
Period & Period^2	✓	✓	✓	✓
Constant	59.23***	60.31***	21.47***	40.74***
	(3.939)	(3.991)	(4.291)	(2.660)
Observations	1419	1419	1338	1338
R^2	0.0427	0.0496	0.644	0.331

Notes: Panel random effects (RE) and fixed effects (FE) regressions on wages for the BASE and the CasRP treatment. Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

wages are more rigid in recession in the CasRP treatment than in the BASE treatment. Because for both specifications the R^2 is of course very low, regression (3) adds some history data about last periods' wages and effort levels either for the case a recession has occurred last period or has not occurred. Additionally, *first wage* adds the very first wage firms choose without having any interactions with other participants. It might reflect to which extent each participant is a selfish or a fair-minded player. These variables sharply increase the R^2. The main results of regression (2) remain, however, unchanged. Regression (4), using fixed effects as a control instead of random effects, confirms these results. Hence in combination with the results of non-parametric tests presented earlier, I get the following result:

> ***Result 2:*** Wages are more rigid in the CasRP than in the BASE treatment. This supports the CasRP hypothesis, and it questions the social preferences hypothesis.

Regression (3) provides another important insight. The CasRP-dummy is significantly negative, suggesting that firms at least slightly adjust out-of-recession wages downward in the CasRP treatment (compared to the BASE treatment). That is, firms behavior has some kind of implicit-contract flavor: They provide workers with more stable wages but adjust the initial out-of-recession wage downward to compensate for their losses in

Table 3.4: Average Effort: BASE vs. CasRP

	Effort		Effort Cut
	No-Recession	Recession	
BASE	4.0	3.3	**0.7**
CasRP	4.1	3.8	**0.3**

recession. This result is not confirmed by regressions (1)-(2) and cannot be confirmed by regression (4). Judged by regression (3)'s R^2, however, this regression seems to be more valid than the very simple regressions (1)-(2). Hence:

> ***Result 3:*** In the CasRP treatment, firms behavior has an implicit-contract flavor: Out of recession, wages are slightly smaller in the CasRP treatment than in the BASE treatment. Firms, however, compensate workers for this loss by not cutting wages in recession as much as in the BASE treatment.

Finally, we know that wage cuts are smaller in the CasRP than in the BASE treatment. Additionally however, out of recession wages are already smaller in the CasRP than in the BASE treatment. So are wages in recession higher in the CasRP than in the BASE treatment, as the non-parametric test (on the individual level) suggested? They are but only with marginal significance (testing for regression (3): Z = 1.71, two-sided $p = 0.087$). Overall, the question that arises from these results is what drives the treatment differences.

3.4.2 Workers' Behavior and Treatment Differences

In this section, I will analyze to what extent differences in workers' behavior can explain differences in the firms' wage-setting behavior, observed in the previous section. Table 3.4 provides a first average measure of workers' behavior in different treatments. We see that in both treatments average effort is cut. For the BASE treatment, this cut is significant according to one-sided Wilcoxon signed rank tests (Session: $p = 0.014$, Individual: $p = 0.010$). For the CasRP treatment, statistical results whether effort is reduced or not are mixed (*Direct* effort cut - Session: $p = 0.014$, Individual: p = 0.027; *Indirect* effort cut - Session: $p = 0.124$, Individual: $p = 0.171$). The effort cut in the CasRP is lower than in the BASE treatment, but this difference is only marginally significant according to one measure[31] But even if there was a difference between treatments, this might of course just be a reflection of lower wages paid by firms in the BASE treatment.

Figure 3.3 gives a first impression why we might observe a treatment difference. This figure reports workers' reaction, their effort change, to different wage changes of firms

[31] One-sided Wilcoxon rank sum test: *Direct* effort cut - Session: $p = 0.075$, Individual: $p = 0.026$; *Indirect* effort cut - Session: $p = 0.261$, Individual: p = 0.225.

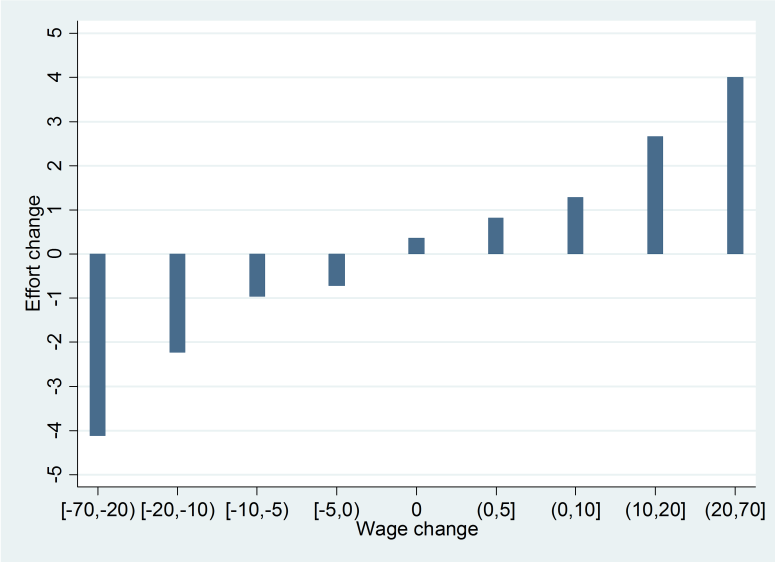


Figure 3.3: Workers' Reaction to Wage Changes - CasRP

for the CasRP treatment (for which *direct* wage cut data is available).[32] If firms e.g. do not change wages at all (wage change of zero), workers react by slightly increasing their effort by roughly 0.3. The overall picture of the graph seems to be that wage cuts are punished and stable (or increased) wages are rewarded. Figure 3.3 does, however, not control for the wage level. It would be possible that the whole pattern just reflects the positive relationship between effort and wage, typically observed in gift-exchange games. Importantly, however, wage cuts seem on average to be punished strong enough to make them unprofitable, whereas stable (or increased) wages are rewarded strong enough to make them profitable. A wage cut of at most 5 is punished by an effort decrease by more than 0.5, whereas a wage increase is also rewarded by an effort increase by more than 0.5. Hence on average, increasing a wage by at most 5 points increases firms' payoff, whereas decreasing a wage by at most 5 points decreases firms payoffs. This punishment-reward pattern of workers' behavior provides a first hint that a reference-point effect might matter in the CasRP treatment.

The regression analysis of Table 3.5 confirms this conjecture. Table 3.5 only uses data from the CasRP treatment and provides panel fixed effects (FE)[33] regressions. This time, the dependent variable is effort. Regression (1) just explains effort by wages, a quadratic wage term, and a constant, controlling for a quadratic time trend. Regression (2) includes a recession dummy that is significant. Workers effort in the CasRP treatment is 0.29 points higher in recession than out of recession. So, although on average less effort is provided in recession as shown by Table 3.4, controlling for wages, workers provide slightly more effort.

[32]For the BASE treatment, direct wage data is by design not available and hence we cannot report results similar to Figure 3.3 for the BASE treatment.

[33]Because I do not want to include time-invariant variables in these regressions, no random effects (RE) regressions are provided. These regression, however, would only lead to results with even slightly higher levels of significance.

Table 3.5: Panel Regressions on Effort - CasRP

	(1) FE	(2) FE	(3) FE	(4) FE
Wage	0.109**	0.113**	0.113**	
	(0.0316)	(0.0308)	(0.0313)	
Recession-dummy		0.285*	0.408**	0.379*
		(0.132)	(0.150)	(0.150)
Rec x Wage cut			-0.257**	-0.176**
			(0.0852)	(0.0648)
Wage^2	✓	✓	✓	
Wage-dummies				✓
Period & Period^2	✓	✓	✓	✓
Constant	-2.532	-2.710**	-2.695**	8.371***
	(0.875)	(0.821)	(0.830)	(0.687)
Observations	827	827	827	827
R^2	0.669	0.671	0.672	0.682

Notes: Panel fixed effects (FE) regressions on effort, only for the CasRP treatment. Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Regression (3) adds an additional dummy (*Rec* × *Wage cut*) that only has a value of one if a recession occurs and the firm cuts wages. In this specification, the recession dummy reflects about how much workers increase their effort when a recession occurs and wages are kept stable (or are even increased). Hence, if a recession occurs and wages are not cut, average effort is about 0.41 points higher than average effort out of recession. If wages, however, are instead cut, average effort is 0.26 points lower than if wages are at least kept stable, and this difference is significant at the 5%-level. Importantly, the effort increase for wage cuts (0.15 points) is statistically not different from zero (Z = 1.14, two-sided $p = 0.256$). Overall this means that, controlling for the wage level, workers only exert more effort in recession for stable (or increased) wages but punish wage cuts relative to wage increases by not increasing effort in recession. This suggests that subjects have established a reference point that leads them to relatively punish wage cuts beyond the pure effect of a lower wage level that is implied by a wage cut.

A potential concern could be that the wage-cut dummy does not really reflect a reference point but captures non-linearities in the wage-effort relationship. Wages that are cut are on average lower than stable (or increased) wages. In order to control for this problem, a quadratic wage term is included in the regressions (1)-(3). Regression (4) provides a different control by including wage-dummies for wage intervals with length of 5 instead of a quadratic wage term.[34] Results change slightly but remain significant at least at the marginal level.[35] Hence, overall, we have the following result:

[34]Overall, 14 wage-dummies are included: $[30, 35], (35, 40], \ldots, (95, 100]$.

[35]Additionally, Table B.1 (Appendix B.1) provides a similar analysis as Table 3.5 but focuses only on recession data. The reason also to use only recession data is that in these regressions wages, the quadratic wage term, and the wage dummies capture precisely the wage-effort relationship in recession and not the overall relationship in and out of recession, as in Table 3.5. Results of Table B.1 confirm those of Table

Table 3.6: Panel Regressions on Effort - BASE and CasRP

	(1) RE	(2) RE	(3) RE	(4) RE	(5) FE	(6) FE
Wage	0.135***	0.137***	0.137***	0.163***	0.135***	0.162***
	(0.0299)	(0.0297)	(0.0301)	(0.0339)	(0.0301)	(0.0334)
CasRP-dummy	0.225	0.341	0.340	1.164		
	(0.288)	(0.266)	(0.267)	(1.149)		
Recession-dummy	0.422***	0.637***	0.634***	0.556***	0.628***	0.547***
	(0.110)	(0.145)	(0.145)	(0.174)	(0.139)	(0.168)
Rec x CasRP		-0.386***				
		(0.148)				
Rec x CasRP x Wage cut			-0.550***	-0.402*	-0.548***	-0.397*
			(0.147)	(0.216)	(0.142)	(0.212)
Rec x CasRP x No wage cut			-0.228	-0.150	-0.220	-0.139
			(0.176)	(0.226)	(0.172)	(0.222)
Wage*CasRP				-0.0501		-0.0494
				(0.0465)		(0.0462)
$Wage^2$	✓	✓	✓	✓	✓	✓
Wage*CasRP & $Wage^2$*CasRP				✓		✓
Period & $Period^2$	✓	✓	✓	✓	✓	✓
Constant	-3.281***	-3.413***	-3.401***	-3.831***	-3.169***	-3.123***
	(0.685)	(0.680)	(0.690)	(0.853)	(0.732)	(0.624)
Observations	1419	1419	1419	1419	1419	1419
R^2	0.526	0.528	0.528	0.531	0.622	0.628

Notes: Panel random effects (RE) and fixed effects (FE) regressions on effort for the BASE and the CasRP treatment. Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

> ***Result 4:*** In the CasRP treatment, workers react differently to wage cuts compared to stable or increased wages even if we control for the wage level. This provides evidence that a reference-point effect matters in the CasRP treatment. Wage cuts are punished beyond the pure reduction of wages compared to stable or increased wages.

Having found evidence in favor of a reference-point effect in the CasRP treatment, Table 3.6 shows to what extent this effect leads to a treatment difference.[36] It provides panel random effects (RE) and fixed effects (FE) regressions with effort as the dependent variable for both main treatments. Regression (1) explains effort by wages, a quadratic wage term, a treatment dummy (CasRP dummy), and a recession dummy, controlling for time trends. In contrast to the wage regressions, workers seem not to behave differently per se in the two different treatments. The *CasRP-dummy* is insignificant for all specifications. In regression (1), the recession dummy is significant, meaning that over both treatments average effort is roughly 0.42 points higher in recession than out of recession. Regression (2) adds an interaction dummy ($Rec \times CasRP$) between the recession dummy and the treatment dummy. This dummy shows that the effort increase in recession in the CasRP

3.5 but are in general statistically less significant. The number of observations, however, reduces sharply. Still, the wage-cut dummy is (at least marginally) significant in five out of six specifications.

[36] See Kocher and Strasser (2011, p. 19) for a similar analysis.

treatment is 0.39 points lower than in the BASE treatment. Regression (3) then shows why we observe a difference between the two treatments by splitting the interaction dummy of regression (2) into two separate dummies. The first dummy (*Rec* × *CasRP* × *Wage cut*) measures the effort reaction of workers in the CasRP treatment in recession when their wage has been cut, whereas the second dummy (*Rec* × *CasRP* × *No wage cut*) measures this reaction for stable (or increased) wages. The results are illuminating: In the BASE treatment, average effort in recession increases significantly by 0.63 points. This increase does not change significantly if we look at the CasRP treatment for stable (or increased) wages. If we look, however, at wage cuts in the CasRP treatment, we observe that the increase in effort due to a recession is 0.55 points lower than in the BASE treatment. This difference is significant at the 1% level. In other words, there is no difference in workers' reaction in the BASE and the CasRP treatment if we consider stable or increased wages (controlling for the wage level), but there is a difference between the two treatments if we consider wage cuts.[37] Controlling for the wage level, wage cuts are punished in the CasRP treatment also compared to the BASE treatment and not only compared to stable (or increased) wages as seen in Table 3.5. Importantly, this result remains at least marginally significant if I include an additional interaction term between wages and the CasRP treatment (*Wage*CasRP*), as shown in regression (4). Additionally, results shown so far are basically fully confirmed by using fixed instead of random effects in regressions (5) and (6). Summing up the results of Table 3.5 and Table 3.6, gives the following result:

> ***Result 5:*** In both treatments, workers exert on average more effort in recession when controlling for the wage level. Overall, however, this effort increase in recession is insufficient for sharing the profit shock with the firm. More importantly, in the CasRP treatment, this effort increase is only found for stable (or increased) wages. Wage cuts are punished by not increasing effort. Hence, there is a reference-point effect in the CasRP treatment and this effect leads to significant differences in workers' reactions between treatments.

What role does it play that firms receive only limited information about workers' reaction to wage changes in the CasRP (and in the BASE) treatment? Importantly, I find evidence that there is a reference-point effect in the CasRP treatment, but this effect does not lead to overwhelming payoff differences of firms. Using the coefficients of regression (3) of Table 3.5, cutting wages seems not to be a profitable strategy for firms, but differences are not too big. A firm, paying a wage of $w = 60$, that does not cut the wage in recession, will on average receive an effort level of about 4.5, leading to an average payoff of 15 in

[37] Unsurprisingly, the *recession-dummy* and the *Rec* × *CasRP* × *Wage cut*- dummy are together statistically not different from zero (Z = 0.71, two-sided $p = 0.478$) whereas the *recession-dummy* and the *Rec* × *CasRP* × *No Wage cut*- dummy are statistically different from zero (Z = 2.97, two-sided $p = 0.003$). Hence, in the CasRP, relative effort increases for stable (or increased) wages, but it does not increase for wage cuts.

Table 3.7: Average Wages: BASE vs. CasRP/ WasRP

	Wage		Wage Cut	Relative
	No-Recession	Recession		Wage Cut
BASE	60.1	47.0	**13.1**	0.21
CasRP	57.9	51.9	**6.0**	0.10
WasRP	60.4	53.8	**6.6**	0.11

recession. A firm cutting wages from $w = 70$ to $w^{Rec} = 60$ will receive an average payoff of 12.5. Hence, effects are fairly small. An additional payoff analysis (Appendix B.3) confirms that firms cutting wages earn less than firms paying stable (or increased) wages in recession. But also in this analysis differences are not too large. So, on the one hand it seems more likely that with more information firms may pay more rigid wages because they might not have been able to fully detect workers' punishment behavior in the CasPR treatment. On the other hand, I cannot fully exclude that some firms may also pay more flexible wages because they might have overestimated workers' punishment behavior. The CasRP-F, discussed in the next section, shows whether more information leads to more rigid wages or not.

3.4.3 Control Treatments

This section presents the results of the two control treatments: First, the WasRP treatment is presented. Afterwards, the results for the CasRP-F treatment are outlined.

WasRP

In my first control treatment, I analyze through which channel contracts serve as reference points. Is an explicit conclusion of a contract between firms and workers necessary to create resistance against wage cuts, or is knowing the payoff-irrelevant out-of-recession wage (and that it has been adjusted downward) already enough because workers still perceive the situation *as if* a contract had been concluded? Basically, if wage cuts of similar magnitude are found in the WasRP treatment as in CasRP treatment, this indicates that contracts serve as reference points because they determine the out-of-recession wage and convey this wage to workers. If, however, wages are as flexible in the WasRP treatment as in the BASE treatment, this indicates that contracts serve as reference points in the sense that additionally an explicit contract conclusion is necessary for generating wage rigidity.

Table 3.7 shows average wages and the two wage cut measures of the two main treatments and the WasRP treatment. Although the wage cut in the WasRP treatment is slightly higher than in the CasRP treatment, it is still roughly of the same size. Additionally, non-parametric tests do not show a significant difference between the CasRP and the

Table 3.8: Panel Regressions on Effort - WasRP

	(1) FE	(2) FE	(3) FE	(4) FE
Wage	0.104**	0.113***	0.113***	
	(0.0274)	(0.0267)	(0.0243)	
Recession-dummy		0.454**	0.739**	0.702**
		(0.176)	(0.189)	(0.182)
Rec x Wage cut			-0.608***	-0.495***
			(0.139)	(0.113)
Wage^2	✓	✓	✓	
Wage-dummies				✓
Period & Period^2	✓	✓	✓	✓
Constant	-1.637	-2.094*	-1.982*	7.379***
	(0.884)	(0.889)	(0.826)	(0.639)
Observations	643	643	643	643
R^2	0.562	0.571	0.576	0.588

Notes: Panel fixed effects (FE) regressions on effort, only for the WasRP treatment. Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

WasRP treatment[38], whereas there is a significant difference between the BASE and the WasRP treatment.[39] This finding is also supported by the regression analysis of Table B.2 (Appendix B.1). It shows that in the WasRP treatment wages are adjusted downward but to a lesser extent than in the BASE treatment, as in the CasRP treatment.

Table 3.8 supports the view that the CasRP and the WasRP treatment lead to similar results. It provides the same analysis for the WasRP treatment that Table 3.5 provides for the CasRP treatment and shows that there is a similar reference-point effect in the WasRP treatment as observed in the CasRP treatment. Wage cuts are punished relative to stable (or increased) wages. Table B.3 (Appendix B.1) shows that this reference-point effect leads to a treatment difference in workers' behavior between the BASE and the WasRP treatment, as already observed for the CasRP treatment. Overall, this gives the following result:

> ***Result 6:*** Contracts already serve as reference points in my setting because they determine the out-of-recession wage and convey this information to workers. An explicit contract conclusion is surprisingly not necessary to generate wage rigidity. In the WasRP treatment, workers might perceive the situation *as if* a contract had been concluded.

[38]CasRP vs. WasRP - one-sided Wilcoxon rank-sum test: For the (raw) *wage cut* measure: *Direct* wage cut - Session: $p = 0.212$, Individual: $p = 0.191$; *Indirect* wage cut - Session: $p = 0.436$, Individual: $p = 0.407$. For the *relative wage cut* measure: *Direct* - Session: $p = 0.212$, Individual: $p = 0.176$; *Indirect* - Session: $p = 0.436$, Individual - $p = 0.342$.

[39]Base vs. WasRP - one-sided Wilcoxon rank-sum test: For the (raw) *wage cut* measure: *Direct* wage cut - Session: $p = 0.003$, Individual: $p = 0.039$; *Indirect* wage cut - Session: $p = 0.027$, Individual: $p = 0.047$. For the *relative wage cut* measure: *Direct* - Session: $p = 0.003$, Individual: $p = 0.039$; *Indirect* - Session: $p = 0.013$, Individual - $p = 0.072$.

Table 3.9: Average Wages: BASE vs. CasRP/ CasRP-F

	Wage		Wage Cut	Relative
	No-Recession	Recession		Wage Cut
BASE	60.1	47.0	**13.1**	0.21
CasRP	57.9	51.9	**6.0**	0.10
CasRP-F	58.8	53.4	**5.4**	0.09

I interpret this result as a indication how strong the behavioral forces are that lead to the reference dependence of fairness considerations. A violation of an explicit agreement between workers and firms is not necessary to create resistance against wage rigidity. Knowing the personal out-of-recession wage, workers would have gotten if a recession had not occurred, is already enough to invoke feelings of entitlement that lead to wage rigidity.

CasRP-F

Finally, I analyze the impact of additional feedback for firms. As outlined, there might be an information problem for firms in the CasRP treatment, potentially prohibiting even more rigid wages. So, does more information lead to more rigid or even completely rigid wages in the CasRP-F treatment? Table 3.9 provides the answer and presents average wages and the two (*indirect*) wage cut measures for all three treatments.[40] Obviously, the wage cut in the CasRP-F treatment is - as expected - slightly smaller than in CasRP treatment, but it is still roughly of the same size. Differences between the CasRP and the CasRP-F treatment are in general insignificant according non-parametric tests[41], whereas differences between the CasRP-F and the BASE treament are unsurprisingly significant.[42] Table B.2 corroborates this finding by providing a regression analysis similar to those of Table 3.3. There is no difference in recession wages between the CasRP and the CasRP-F treatment (Regression 3 - *Rec* × *CasRP-F* – *Rec* × *CasRP*: Z = 0.54, two-sided $p = 0.587$). Hence, at least on average, additional feedback for firms seems not to alter the results and does not lead to fully rigid wages.

The average results, however, mask substantial differences between the CasRP and the CasRP-F treatment. Figure B.1 (Appendix B.1) shows a histogram of wage changes for

[40]The direct wage cut measures for the CasRP and the CasRP-F are the following: For the CasRP-F, the direct no-recession wage is 57.1, the direct wage cut: 3.6, the direct relative wage cut: 0.06. For the CasRP, the direct no-recession wage is 57.0 the direct wage cut is 5.1 and the direct relative wage cut is 0.09.

[41]CasRP vs. CasRP-F - one-sided Wilcoxon rank-sum tests - : For the (raw) *wage cut* measure: *Direct* wage cut - Session: $p = 0.212$, Individual: $p = 0.071$; *Indirect* wage cut - Session: $p = 0.261$, Individual - $p = 0.255$. For the *relative wage cut* measure: *Direct* - Session: $p = 0.212$, Individual: $p = 0.247$; *Indirect* - Session: $p = 0.316$, Individual: $p = 0.286$.

[42]Base vs. CasRP-F - one-sided Wilcoxon rank-sum test: For the (raw) *wage Cut* measure: *Direct* wage cut - Session: $p = 0.027$, Individual: $p = 0.002$; *Indirect* wage cut - Session: $p = 0.027$, Individual - $p = 0.013$. For the *relative wage cut* measure: *Direct* - Session: $p = 0.004$, Individual: $p = 0.003$; *Indirect* - Session: $p = 0.019$, Individual: $p = 0.014$.

Table 3.10: Panel Regressions on Effort - CasRP-F

	(1) FE	(2) FE	(3) FE	(4) FE
Wage	0.131***	0.131***	0.131***	
	(0.0162)	(0.0179)	(0.0182)	
Recession-dummy		0.0225	0.0828	0.0993
		(0.205)	(0.248)	(0.245)
Rec x Wage cut			-0.160	-0.170
			(0.218)	(0.233)
Wage^2	✓	✓	✓	
Wage-dummies				✓
Period & Period^2	✓	✓	✓	✓
Constant	-2.750***	-2.756***	-2.725***	8.206***
	(0.572)	(0.673)	(0.670)	(0.510)
Observations	796	796	796	796
R^2	0.669	0.669	0.670	0.680

Notes: Panel fixed effects (FE) regressions on effort, only for the CasRP-F treatment. Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

both treatments. This figure suggests that stable (or increased) wages are slightly more likely in the CasRP-F treatment, as predicted by Hypothesis 3. This effect is, however, negated by a sharp increase in the very strong wage cuts (above 20 points).[43] So, more information leads to more rigid wages to some extent, but it also increases very large wage cuts, leading to an average result comparable to the CasRP treatment.[44]

How can we explain this increase in variation in the CasRP-F treatment? The increase in the very large wage cuts can potentially not be explained by firms that overestimate workers' punishment in the CasRP treatment. Even with more information, those firms should only adjust wages downward by about 10 points. It seems that also workers change their behavior in the CasRP-F treatment compared to the CasRP treatment. Table 3.10 presents regressions analogous to the regressions of Table 3.5 for the CasRP-F treatment. Unlike before, these regressions show that workers surprisingly do not (significantly) increase their effort level in recession in the CasRP-F treatment, even when wages are increased or kept stable. There is also no significant difference between wage cuts and stable (or increased) wages. Hence, a significant reference-point effect, as observed in the CasRP and the WasRP treatment, cannot be found for the CasRP-F treatment. I do, however, find a different kind of reference-point effect compared to the BASE treatment:

[43]Overall, in the CasRP-F treatment 60.3% of all wages are increased or kept stable whereas this is only the case for 51.2% in the CasRP treatment and this difference is marginally significant (Z = 1.84, two-sided $p = 0.065$). In the CasRP-F treatment, 6.5% of all wages are cut by more than 20 points whereas this is only the case for 2.9% of wage changes in the CasRP treatment. This difference is again marinially significant (Z = 1.72, two-sided $p = 0.083$).

[44]These observations are also in line with the fact that in Table B.2 (Appendix B.1) the standard error of the *Rec* × *CasRP-F*-dummy is higher than the standard error of the *Rec* × *CasRP*-dummy. The variation in wage cuts seems to be higher in the CasRP-F treatment due to a higher frequency of the very large wage cuts and the very strong wage increases.

Controlling for the wage level, workers do not increase their effort in recession and this fact holds independent of whether the wage has been cut or not (which is also supported by Table B.3). This may provide a potential explanation for firms' behavior: Wage cuts are punished as in the CasRP treatment by not increasing effort (when controlling for wages). Hence, many firms pay more rigid wages compared to the BASE treatment. Additionally, however, workers do not reward paying stable (or increased wages) by providing relatively more effort. Hence, at least some firms seem to react by cutting even high wages to the minimum wage, potentially because they want to punish non-reciprocal workers. Overall, this gives the following result:

> ***Result 7:*** Additional information does on average not lead to more rigid or completely rigid wages. Hypothesis 3 cannot be confirmed. Nonetheless, average results mask important differences between the CasRP and the CasRP-F treatment. In the CasRP-F treatment, the percentage of stable or increased wages is higher than in the CasRP treatment but also the frequency of very high wage cuts increases. This is potentially due to a change of workers' behavior in the CasRP-F treatment. They do not reward stable (or increased) wages.

Overall, it seems that the intuition of the CasRP-F treatment is right, with more information there is also a slight tendency to pay more rigid wages because firms are better informed about workers' behavior. The change in the information condition, however, seems to also alter workers' behavior. Why workers behave differently is a matter of speculation. A plausible suggestion might be that with better information workers realize that they a have powerful punishment device. With this knowledge, workers, however, seem to forget that they also have to reward stable (or increased) wages, leading to the result that some firms cut wages sharply. Hence, the result that more information does not lead on average to more rigid wages seems at least partially to be an artifact of my experimental design.

3.5 Conclusion

Are fairness considerations reference-point dependent and does this dependence provide an explanation for wage rigidity? And do contracts initially concluded before recessions serve as reference-points? In my experimental design, I mainly compare two different situations: In one treatment, a contract is initially concluded before recessions and has to be renegotiated if a recession occurs. In this treatment, initial contract conditions can serve as a reference point. Workers may feel entitled to the initial wage and hence punish wage cuts. In another treatment, such initial contracts do not exist, and contracts are only concluded after both contract partners know whether a recession has occurred or not.

My main findings are the following: In both treatments - with or without initial contracts - wages are cut in recession. Hence, I do not observe complete wage rigidity in both treatments. Importantly, however, wages are significantly more rigid with initially concluded contracts compared to a situation where no such initial contracts exist. Additionally, there is clear evidence that initially concluded contracts serve as reference points. Workers in this treatment seem to feel entitled to the initial wage and punish wage cuts compared to stable (or increased) wages. So fairness considerations are reference-point dependent and contracts serve as reference points. Importantly, the overall reference-point effect seems to be surprisingly strong if we bear in mind that all my results hold even though subjects do not get used to their wage over a longer period of time and wage cuts in the laboratory do not put subjects' standard of living at risk, as they potentially do in the field. Without reference points, subjects seem to be well aware that moderate wage cuts are fair, but introducing the possibility of reference points seems to completely overrule these considerations.

An additional control treatment shows through which channel contracts serve as reference points. Providing the personal out-of-recession wage also in recession already is sufficient to make fairness considerations reference-dependent. Hence, contracts serve as reference points by just determining the out-of-recession wage and conveying this information to workers. Behavioral forces behind these reference points are obviously strong enough to induce wage rigidity even without an explicit contract conclusion. Workers' potential perception *as if* a contract had been concluded seems to be sufficient to create reference points that lead to wage rigidity. Finally, offering firms more information about workers' behavior does surprisingly not lead to more rigid wages, at least on average.

Hence overall, this paper provides experimental evidence that the dependence of fairness considerations on reference points provides one explanation for wage rigidity. A crucial question for future research is to what extent such a reference dependence is also valid for booms. Do initially concluded contracts constrain workers from asking for higher wages if firms make higher profits because of good business conditions? Or, do we observe an asymmetry between booms and busts? If the underlying reason for workers' resistance against wage cuts can be found in some form of self-serving bias, reference points might not drive workers' behavior if the consequence would be a foregone higher wage in a boom.

Chapter 4

The Winner's Curse: Contingent Reasoning & Belief Formation[1]

4.1 Introduction

In recent research, it has been claimed that people have difficulties in performing contingent reasoning on hypothetical events. This form of cognitive limitation might be relevant in many strategic situations where people must think about hypothetical situations. Charness and Levin (2009) claim that decision makers' problems to condition on the crucial future event of winning the auction is at the origin of the winner's curse - the phenomenon that people in common value auctions systematically overbid which typically results in severe losses. In addition, Esponda and Vespa (2014) come to a very similar conclusion in an experimental voting environment. The authors analyze a voting setting in which strategic voters have to condition on the hypothetical event that their vote is pivotal, given their beliefs about the equilibrium strategies of the other players. They observe that subjects have severe difficulties in drawing inferences from hypothetical events.

In this paper, we will focus on the winner's curse (WC). In a common-value auction (CVA) setting, decision makers have at least two interrelated problems in finding the equilibrium strategy. First, they have to act strategically in the sense that they have to form beliefs about their opponents' bidding strategy and it is unclear to them to what extent their opponents are rational. Second, if they have formed these beliefs, they have to condition their actions on the critical future event of winning. Charness and Levin (2009) claim that the second step is crucial in understanding why subjects fall prey to the WC. The authors substantiate their claim by observing the WC in an individual choice setting (acquiring-a-company game with computer opponents) in which subjects do not have to form beliefs about their opponents. Even without the additional problem of strategic interaction with other players, subjects seem to have difficulties in avoiding the WC in an

[1]This chapter is based on joint work with Stefan Penczynski.

environment that is characterized by a similar adverse selection problem than the CVA setting.

But can we really infer from the analysis of Charness and Levin (2009) that the cognitive difficulty of performing contingent reasoning on hypothetical events is at the origin of the WC in CVA setting? Two central problems arise that have also been outlined by Ivanov et al. (2010, p. 2). First, in general, the acquiring-a-company game is not a CVA even if we - unlike Charness and Levin (2009) - allow for strategic interaction with human opponents. The game mimics the lemons market setting of Akerlof (1970). Although the structure of both games is fairly comparable, it is obviously not a priori clear that an observation in one setting extends to the other setting. Second, and more importantly, observations made in an individual-choice problem might not necessarily extend to strategic settings. Cognitive strategies applied in both settings could deviate from each other. Moreover, even if conditional reasoning on hypothetical events is a problem both in the individual acquiring-a-company setting and in the CVA setting, it is unclear to what extent an additional problem with belief formation in a typical CVA setting with human opponents provides another obstacle to avoid the WC. Ivanov et al. (2010) provide indirect evidence against models that try to explain the WC by maintaining the best-response assumption but allowing for inconsistent beliefs (*cursed equilibrium* by Eyster and Rabin 2005 and *level-k* by Crawford and Iriberri 2007). In the experimental setting of Ivanov et al. (2010), subjects seem not to form beliefs in a way these belief-based models suggest, potentially they do not form beliefs at all.

Overall, the main question this paper tries to answer is how important the cognitive limitation of not being able to perform contingent reasoning on hypothetical events is in a typical CVA setting with human opponents. In other words, if we switch the conditioning problem off, will decision makers be able to avoid the WC? Or does the problem of belief formation provide another obstacle to understand the game and avoid the WC even without the conditioning problem? We try to shed light on these questions with the help of the following experimental design: We compare decision makers' behavior in a simplified CVA setting with their behavior in a transformed game that approximates the structure of the CVA game but which does not require contingent reasoning on hypothetical events. Our starting point is a typical first-price CVA game in which subjects receive randomly determined information signals around a common value. It has been shown that in such a setting subjects usually fall prey to the WC. We simplify this environment by only allowing for two kinds of information signals and by restricting the number of bidders to two. The purpose of this simplification is the following: Reducing the complexity of the auction setting makes it possible to construct a number guessing game with rules that make this game strategically very similar to the auction setting but that eliminate the need to reason about hypothetical events.

We then implement two different treatments: In the *AuctionFirst* treatment, subjects

first play the simplified CVA game and afterwards play the transformed game. In the *TransformedFirst* treatment, this sequence of events is reversed. Our design allows us to make a between-subject as well as a within-subject analysis. Besides analyzing the conditioning problem, we want to explicitly test whether belief formation is another problem for avoiding the WC. For this purpose, for both treatments, we implement for both the auction game and for the transformed game a second stage in which subjects play against a computer opponent with a known strategy. Hence, in this second stage, the belief-formation problem is basically switched off. If subjects are able to best respond in the stage with a computerized opponent, but are unable to find an equilibrium in the stage with human opponents, the cognitive problems associated with belief formation seem to provide an additional obstacle in avoiding the WC.

We have three main results: First, subjects are to a larger extent able to avoid the WC in the transformed game than in the CVA game (for both human and computerized opponents). Hence, the claim of Charness and Levin (2009) that decision makers' difficulty with contingent reasoning on hypothetical events is at the origin of the WC is confirmed also in a CVA setting with human opponents. Second, the authors claim has, however, to be qualified to the extent that contingent reasoning is not the only obstacle in avoiding the WC. Participants seem to understand both games better without strategic interaction with human opponents and avoid the WC to a larger degree in the setting with computerized opponents. Importantly, strategic uncertainty alone seems to be unable to explain the observed differences. Many subjects play strategies in the games with human opponents that cannot be best response if subjects are uncertain about other subjects behavior. From this observation, we infer that subjects do not only have problems to form correct beliefs, but belief formation per se seems to be an additional problem for avoiding the WC. And in our setting, this obstacle is of similar magnitude than the conditioning problem. Finally, we observe that playing the transformed game first helps subjects to understand the original CVA game whereas the reverse is not true. This suggest that contingent reasoning on hypothetical events indeed is an obstacle to subjects' understanding of the game. Subjects only learn between games when the game in which such kind of contingent reasoning is not necessary is played first.

Starting with Bazerman and Samuelson (1983) as well as Kagel and Levin (1986), numerous experimental papers have shown that experimental subjects often fall prey to the WC and our basic experimental design follows the experimental setup of Kagel and Levin (1986). The WC has also been shown to be valid in different common-value settings, see e.g. Avery and Kagel (1997), Goeree and Offerman (2002), Ivanov et al. (2010). Additionally, authors that explicitly have tried to test the robustness of the WC have mostly found support for the WC, see e.g. Lind and Plott (1991) and Grosskopf et al. (2007). Besides the paper of Charness and Levin (2009) (and to some extent also Esponda and Vespa 2014), the work most closely related to our work seems to be: Levin and Reiss (2012). The

authors construct a behavioral auction design in which the payment rule incorporates the adverse selection problem that is at the origin of the WC. Importantly, the authors only adjust the payment rule but do not change the general structure of the auction setting as we do in our setting. Additionally, in their setting, the authors observe that the WC is still fairly prominently present in their data. Overall, the literature has failed so far to provide evidence that contingent reasoning is at origin of the WC also in a CVA setting with human opponents. This chapter provides this evidence. The rest of this chapter is organized as follows: Section 4.2 describes the experimental design and our hypotheses. Section 4.3 provides our experimental results and Section 4.4 concludes.

4.2 Experimental Design and Hypotheses

In this section, we first outline the games used in our experiment (Section 4.2.1). Afterwards, the experimental design/treatments and the procedures are discussed (Section 4.2.2). Finally, our hypotheses are presented (Section 4.2.3).

4.2.1 The Games

In our experimental design, we will use two different games: a (simplified) standard *auction game* and a strategically very similar *transformed game* that does not require subjects to condition their decisions on the hypothetical event of winning the auction. The starting point for both games is a standard first-price CVA setting (compare Kagel and Levin 1986): There are n bidders and a common value $W^* \in [\underline{W}, \overline{W}]$ which is the same for each bidder. Each bidder receives a private signal $x_i \in [W^* - \delta, W^* + \delta]$, with $\delta > 0$. Bidders make bids in a sealed-bid first-price auction in which the highest bidder wins the auction and pays his bid: The available actions (absolute bids) are $a_i \in [\underline{W}, \overline{W}]$. A bidder's payoff who makes the higher bid and wins the auction is $u_i = W^* - a_i$ because of the first-price auction setting. In case, a bidder does not make the higher bid, the payoff is $u_i = 0$.

(Simplified) Auction Game

We simplify this general setting mainly by allowing only for two signals and by restricting the number of subjects who bid for the commodity to two. Additionally, the common value W^* is uniformly distributed in $[25, 225]$.[2] Bidders receive a private binary signal $x_i \in \{W^* - 3, W^* + 3\}$ and this signal is drawn without replacement. Hence, bidders know

[2]More precisely, the common value is a random variable that is uniformly distributed in $(45, 205) \subset [25, 225]$. We communicate, however, to subjects that bids are randomly drawn from the interval $[25, 225]$. We do this in order to avoid common values near 25 or 225 for which subjects could draw inferences from their signal about the true value. Receiving e.g. a signal smaller than 25 reveals that the true value of the commodity is surely higher. Nonetheless, we do not deceive our subjects because we do not specify the distribution of the random variable. We only tell subjects that the commodity's value is randomly determined and can never be lower than 25 and never be higher than 225, which obviously is still true.

that the other bidder just receives the other signal. Additionally, we implement that in case both players bid the same amount, the player that has received the lower signal wins the auction.

In order to analyze the structure of this (simplified) auction game, note that strategies can be expressed in distance b_i between the individual signal x_i and the absolute bid a_i: $b_i = a_i - x_i$. For simplicity, we will call these relative bids also just *bids* below and always specify when we talk about absolute bids. Because our setting is a strategy situation, player i's strategy will depend on the other player's strategy, b_j. Intuitively, relative to the other player's bid b_j, three general options emerge for player i: First, player i can (relatively) overbid the other player by at least 6 points in order to ensure that he[3] always wins the auction. In this case of overbidding, player i naturally wins the auction if he receives the higher signal. But when player i (relatively) overbids by 6 or more points, he wins the auction even when receiving the lower signal because the signals differ exactly by 6. Moreover, overbidding exactly by 6 points is already sufficient for winning the auction because of our tie-breaking rule that implements that in a case of a tie the player with the lower signal wins the auction. Second, player i can (relatively) underbid the other player by at least 6 points in order to ensure that he never wins the auction. This case is just symmetric to the first option. Finally, player i can (relatively) underbid or overbid the other player by less than 6 points. Then, he will only win the auction when he has received the higher signal. But player i will not win the auction, if he has received the lower signal. Obviously, in order to maximize the payoff of this option, relatively bidding slightly less than 6 points, $b_i = b_j - 6 + \epsilon$ (with $\epsilon > 0$ and small), is optimal. In summary, we have three general options:

1. *Sure-win* bidding: For $b_i \geq b_j + 6$, player i will always win the auction.

2. *Sure-lose* bidding: For $b_i \leq b_j - 6$, player i will never win the auction.

3. *Win-if-high* bidding: For $b_j - 6 < b_i < b_j + 6$, player i wins if and only if he has received the higher signal.

Which of these three options is optimal depends on the opponent's bid b_j. We can compute the best response function for different values of the opponents strategy b_j following the outlined options above. If the other player bids fairly high values, $b_j \in [3, \overline{b_j}]$[4], with $\overline{b_j} = \overline{W} - x_j$, it is optimal for player i to follow the second option to never win the auction. The reason is that winning the auction would in this case result in losses for sure (or at

[3]For convenience, we will consider player i as male and the opponent j as female.

[4]Importantly, in our analysis, we do not consider the case that for specific signals (near $\underline{W}$ or $\overline{W}$) very low or very high bids might not be possible. We neglect this problem because we will further restrict the action space below. Because of this additional restriction and because of footnote 2, signals in the experiment will never be too close to $\underline{W}$ or $\overline{W}$ such that subjects can not make use of their (additionally restricted) action space.

most a zero payoff) because the opponent already bids at least the commodity's value, even in the case that the opponent has received the lower signal. Hence, the best response is to bid anything that is relatively below the opponent's bid by at least 6 points: $b_i = d$, where $d \in [\underline{b_i}, b_j - 6]$, with $\underline{b_i} = -(x_i - \underline{W})$, leading to a payoff of 0 because of not winning the auction (even in the case of receiving the higher signal).

If the other player bids fairly low values, $b_j \in [\underline{b_j}, -15]$, with $\underline{b_j} = -(x_j - \underline{W})$, it is optimal for player i to relatively overbid the other player (first option) in order to ensure that he wins the auction even if he has received the lower signal. This strategy is profitable because of the fairly low values the other player bids. Hence, even if player i overbids the other player, he makes a profit. The best response is $b_i = b_j + 6$. In this case, player i wins the auction for sure and earns $u_i = W^* - a_i = W^* - (x_i + b_i) = 3 - b_i$ in case of the lower signal and $u_i = W^* - (x_i + b_i) = -3 - b_i$ in case of the higher signal, with 50 percent probability each.[5]

If the other player bids intermediate values $b_j \in [-15, 3]$, overbidding the opponent when having the lower signal is not profitable any longer because for intermediate values of b_j, player i will not earn enough money by this strategy (compared to the third option). Hence, in case of intermediate values of b_j, it is optimal for player i to relatively underbid the opponent by slightly less than 6 points in order to ensure (according to the third option) that he only wins the auction when he has received the higher signal. Hence, the best response function is $b_i = b_j - 6 + \epsilon$, leading to a payoff of $u_i = -3 - b_i$ (higher signal) and $u_i = 0$ (lower signal), with 50 percent probability each.

In order to create a unique equilibrium prediction for the (simplified) auction game (and in order to avoid out-of-equilibrium collusion) we restrict the players' action space in a way that decision makers can only absolutely overbid and underbid their private signal by at most 8 points, $a_i \in [x_i - 8, x_i + 8]$. Hence, they can also only (relatively) overbid or underbid by at most 8 points, $b_i \in [-8, 8]$. In this setting, bidding $b_i = b_j = -8$ is the unique equilibrium. First of all note, that option 2 (sure-lose bidding) is not a feasible strategy for player i when his opponent bids $b_j = -8$ because he will always win the auction when he has the higher signal for any bid b_i. For this reason, we have to compare options 1 (sure-win bidding) and 3 (win-if-high bidding): When opponent j maximally underbids, $b_j = -8$, it is then straight-forward to see that the third option leads to higher payoffs than the first option: Following the third option, player i should relatively underbid

[5] Concerning the precise cutoff value of -15, note that for an opponent's bid of $b_j = -15$, strategies according to the first option (sure-win bidding) and the third option (win-if-high bidding) lead to (nearly) the same result, at least in expectations. For this reason, $b_j = -15$ is the cutoff at which the other strategy becomes more profitable. Relatively overbidding by six points with $b_i = -9$ leads to a payoff of $u_i = -3 - (-9) = 6$ (higher signal) and $u_i = +3 - (-9) = 12$ (lower signal), with 50 percent probability each. Relatively underbidding by slightly less than 6, $b_i = -21 + \epsilon$, however leads to a payoff of $u_i = -3 - (-21 + \epsilon) = 18 - \epsilon$ (higher signal) and $u_i = 0$ (lower signal), with 50 percent probability each. Hence, in expectations and if continuous bidding is possible, both strategies lead in the limit to the same result.

the opponent by slightly less than 6 points. Naturally, because of the restrictions of the action space, player i can at most only underbid by bidding the same, $b_i = -8$. Player j already bids the lowest possible (relative) amount. Player i then only wins the auction if he has received the higher signal, leading to an expected payoff of $Eu_i = \frac{1}{2}(-3 - b_i) = 2.5$. If player i, however, follows the first option and bids $b_i = b_j + 6 = -2$, he will receive an expected payoff of $Eu_i = \frac{1}{2}(-3 - b_i) + \frac{1}{2}(+3 - b_i) = \frac{1}{2}(-1 + 5) = 2$. Overbidding the opponent is not profitable because it leads to losses in case of receiving the higher signal. Hence, bidding the lowest amount possible is an equilibrium strategy because there are no incentives to deviate.

Additionally, subjects always have incentives to deviate from any pair of strategies in which not both subjects bid $b_i = b_j = -8$ because of the following argument: Note first, that with the restriction of the action space, following option 1 (sure-win bidding) can never be a best response. As shown before, already for $b_j = -8$, this kind of overbidding leads to losses when receiving the higher signal. Obviously, these losses even increase when $b_j > -8$. Hence, subjects always have an incentive to underbid their opponent. If the opponent bids high, $b_j \in [3, 8]$, player i should follow option 2 (sure-lose bidding) and underbid by at least 6 points in order to ensure not to win the auction and face losses. If the opponent makes lower bids, $b_j \in (-8, 3)$, and hence positive payoffs are possible, player i should follow option 3 (win-if-high bidding) and underbid his opponent by slightly less than 6 points. Hence, whenever player j bids more than -8, player i has an incentive to underbid his opponent and if player i underbids but still bids more than -8, player j has an incentive to underbid player i. These underbidding incentives only vanish when no underbidding is possible any more, i.e. subjects bid -8.

These equilibrium considerations so far do not take into account that subjects receiving a signal close to 25 or 225 can infer the commodity's real value. This might not only influence those subjects' strategies that receive signals close to 25 or 225, but it could in principle also influence those subjects' strategies that receive signals well within the interval $[25, 225]$. In Appendix C.1 we show, however, that this influence vanishes very quickly and that $b_i = -8$ remains the equilibrium strategy for all realizations of the commodity's value that occur in the experiment. Or in other words, for the signals occurring in the experiment bidding $b_i = -8$ is the equilibrium strategy.

Overall, at least two major problems might be an obstacle for participants to play the equilibrium strategy. In the auction game described so far, participants face strategic uncertainty. They do not know how others behave and to what extent they are rational. They have to form beliefs about what other subjects do. In addition, when calculating their best responses to their beliefs, they have to condition on the critical event of winning the auction. Both problems (forming beliefs and contingent reasoning) may provide an obstacle on their own to avoid the WC curse. We will analyze these obstacles in two ways. Concerning the problem of belief formation, we will let subjects play the described

auction game as above with a slight modification. In the *auction game with computerized opponents*, subjects will play against a computerized opponent. Subjects in this game know what strategy the computerized opponent follows and hence do not face any strategic uncertainty. We implement the following decision rule for computerized opponents: The computer just bids its signal. Hence, its bid is $b_i = 0$. We implement this rule because it reflects a naive strategy that does not account for the problem of conditioning on the critical event of winning and we want to investigate to what extent subjects are able to play the best response against this strategy. Obviously, relatively underbidding the computer by $b_i = -5.99$[6] is the best response against the computer because it ensures winning only when having received the higher signal and does so at the minimum possible price. Winning even when having received the lower signal would naturally result in losses because the computer in this case has received the higher signal and already overbids the commodity's value by bidding the higher signal. We tackle the problem of having difficulties with contingent reasoning on hypothetical events by introducing a transformed game that is strategically very similar to our auction game but does not require contingent reasoning on hypothetical events.

Transformed Game

The general idea of this *transformed game* is to construct a number choosing game that approximates the features of the auction game when strategies are expressed in distance to the signal. The general structure of the game is that subjects choose a number $n_i \in [-8, 8]$, and the rules of winning the game are determined according to the structure of the auction game as described in the three general options subjects face as described above. Constructing the game in this way leads to a strategically similar game compared to the auction game. Importantly, the conditioning on the hypothetical event of winning, however, has not to be done by participants but is already incorporated in the rules of the game. The main strategic difference between the two games is that the auction game is a game with incomplete information whereas the transformed game is not. Nonetheless, for all relevant signals and commodity's values, subjects should follow the same strategy (bidding $b_i = -8$) in both games. In principle, it would have been possible to make the transformed game fully strategically equivalent to the auction game, but this would have been associated with severe other problems, as will be outlined in more detail after describing the transformed game. Overall, the construction of the transformed game allows us to investigate the differences in play that are due to the cognitive activity related to conditioning.

A drawback of our approach so far is that the two games are behaviorally equivalent in the sense that subjects should follow the same bid in equilibrium, but the framing differs

[6] In the experiment, subjects at least have to round their bids to one cent.

between them. Hence, it would not be sure to what extent potential treatment effects reflect differences in contingent reasoning or in framing. In order to avoid a very different framing, we frame our number choosing game as an auction game with special rules. In addition, we let subjects not just choose a number $n_i \in [-8, 8]$ but shift absolute bids to the same levels as in our auction game. In our transformed game, we again have two subjects. This time two potential common values arise. The common value is either W_1^* or W_2^*, where the difference between both values is exactly 6 and W_1^* always denotes the smaller value. The average value of both common values reflects the signal of the auction game. Either the lower value W_1^* finally realizes (comparable to the high signal case in the auction game) or the higher value W_2^* (comparable to the low signal case) finally realizes according to rules outlined below. In order to roughly establish a comparable level of absolute bids as in the auction game, W_1^* is randomly distributed in the following way: $W_1^* \in [25, 219]$ and $W_2^* \in [31, 225]$.[7] Subjects are allowed to absolutely underbid W_1^* by 5 points and absolutely overbid W_2^* by 5 points, $a_i \in [W_1^* - 5, W_2^* + 5]$. (Relative) bids b_i are measured relative to the average of both common values (reflecting the signal in the auction case), $b_i = a_i - \frac{W_1^* + W_2^*}{2}$, hence we again have $b_i \in [-8, 8]$, as in the auction setting. Which of the two common value arises for which player, depends on the bids of the two players and on chance, exactly following the structure of the (simplified) auction game. In order to establish a good approximation of the auction game by the transformed game, the rules of the transformed game basically follow the three options of the auction game. Although these rules lead to two games, in which subjects should in equilibrium behave identically (for the signals we are interested in), they already contain the conditioning on the hypothetical event of winning which in the auction game has to be done by subjects themselves. The game has three rules, parallel to the options in the auction game:

1. Parallel to option 1 (sure-win bidding): If one player (relatively) overbids the other player by 6 or more points, this player wins the auction for sure and gets the following payoff: With 50 percent probability $u_i = W_1^* - a_i (= W_1^* - (W_1^* + 3 + b_i) = -3 - b_i)$ (mimicking the high signal case of the auction game) and with 50 percent probability $u_i = W_2^* - a_i (= W_2^* - (W_2^* - 3 + b_i) = +3 - b_i)$ (mimicking the low signal case of the auction game).[8]

2. Parallel to option 2 (sure-lose bidding): If one player (relatively) underbids the other player by 6 or more points, this player does not win the auction and his or her payoff is $u_i = 0$.

3. Parallel to option 3 (win-if-high bidding): If the difference of both players' bids is

[7] For reasons of comparability, in the experiment, we adjusted the real interval appropriately, following the adjustment in the auction game.

[8] As in the auction setting, in the instructions, subjects always have to determine absolute bids and payoffs are only described in terms of absolute bids. For the ease of exposition, we however, also present payoffs in terms of (relative) bid b_i.

smaller than 6 points, either player i or player j wins the auction/game with 50 percent probability. In case player i wins the auction, his payoff is $u_i = W_1^* - a_i (= -3 - b_i)$ (mimicking that in this case the player with the higher signal wins in the original auction setting). In case player i does not win the auction, his payoff is $u_i = 0$.

These rules already incorporate the conditioning on the event of winning. In the third rule, the fact that the realized common value always is the lower value W_1^* captures the idea that in the auction game the player with the higher signal wins the auction if bids deviate by less than 6 points. Overall, it holds that the general design and the rules of the transformed game just make explicit what is implicitly given in the auction game and has to be understood by subjects without any further assistance. Hence, by construction of the game, the equilibrium bid of the transformed game is the same as in the auction game (if we abstract from the boundaries): (relatively) underbidding by $b_i = -8$ leads to an absolute bid of $a_i = W_1^* - 5$.

As already outlined above, it is important to note that the transformed game and the auction game are not strategically equivalent in the sense that the auction game is a game with incomplete information whereas the transformed game is not. In the auction game, players do not know whether their opponent has a higher or a lower signal and hence whether the commodity's value is higher or lower than their own signal. Importantly, if player i gets the signal x_i then he thinks the signal $x_i - 6$ is possible for j, hence he thinks that j thinks that the signal $x_i - 12$ is possible for i, hence he thinks that j thinks that i thinks that $x_i - 18$ is possible for j, and so on. This kind of higher order beliefs are not possible in the transformed game because the commodity's value is the same for both players, it just has two possible realizations. Importantly, we show in Appendix C.1 that bidding -8 remains the equilibrium strategy in the auction game for signals that are not close to the lower boundary. Hence, because the difference of higher order beliefs does not change the equilibrium strategy for signals well within the interval $[25, 225]$, both the auction and the transformed game are behaviorally equivalent in the sense that subjects should make the same bid in equilibrium (for the relevant signals and values of the commodity). Moreover, the two games are outcome equivalent in line with Krishna (2009, p. 180): If both opponents submit the same bids in the two games the same outcome regarding who wins the auction and regarding the price of the commodity results. This outcome equivalence also holds for boundary signals and values of the commodity since also at the boundary the same bids in the two games still result in the same outcome, only the equilibrium prediction changes here. Additionally, we expect that the overwhelming majority of subjects might not consider the outlined higher order beliefs.[9]

[9]In a trial session, we implemented the auction game with a communication design similar to Burchardi and Penczynski (2014). In this setting, groups of two communicate about the bidding decision in a way that subjects have an incentive to share their knowledge about the game. In this trial setting, only a minority of groups forms beliefs about how other groups potentially bid. Importantly, no group is concerned about boundary-signals and their implication for their own bidding strategy.

In principle, it would have been possible to implement the transformed game as a game of incomplete information in which the two commodity's values are different for both players. The problem with this approach would, however, be that these different values would not be meaningful in the sense that they do not matter for the rules of the game because these rules would have to rely on comparing both opponents' bids relative to their own mean value of their two values of the commodity, making these different values superfluous. Naturally and in contrast to this observation, different signals for the two opponents are meaningful in the auction game because of the structure of this game. Overall, this approach potentially would have made such a transformed game with incomplete information look very artificial for participants.[10] Overall, we believe that our transformed game presents the best solution of the trade-off between strategic similarity between the two games and the simplicity and reasonableness of the transformed game. Hence, we would argue that our transformed game best approximates the auction game.

Finally, it is straight-forward to implement a transformed game with computerized opponents as we also do it for the auction game. In the auction game, computerized opponents bid the naive strategy of bidding the private signal. In the transformed game, the private signal, however, is reflected by the average of the two common values $\frac{W_1^*-W_2^*}{2}$. As in the auction game, bidding $b_i = -5.99$ is the subjects' best response to this strategy, leading to an absolute bid of $a_i = W_1^* - 2.99$.

4.2.2 Experimental Design and Treatments

Our experiment consists of two treatments: the *AuctionFirst* and the *TransformedFirst* treatment. The general structure of the two treatments is the following: In the *AuctionFirst* treatment, subjects first play the auction game and afterwards the transformed game. In the *TransformedFirst* treatment, this sequence of events is reversed. More precisely, in the *AuctionFirst* treatment, first of all, the general instructions and the instructions for the auction game are read aloud. In these instructions, subjects are informed that they will first make all their decisions in the experiment before receiving any feedback. The experiment is then divided into two parts. In Part I, subjects play three periods[11] of the auction game with human opponents and stranger matching. Afterwards, they play three additional periods of this game with computerized opponents. As outlined before,

[10] Additionally, the rules of the transformed game with incomplete information could also have allowed for the fact that subjects receiving a signal close to 25 know the real commodity's value in the auction game (cf. Appendix C.1). This would have increased the complexity of these rules quite a lot. Additionally, by incorporating this problem in the rules of a transformed game with incomplete information, this problem would have been highlighted quite a lot in this game although it might not have been realized by most people in the auction game because it is in some sense only given implicitly in this game.

[11] Because subjects do not receive any feedback after playing one period, in principle, it would have been possible to just implement one period per game. We wanted, however, to implement more than one period in order to avoid that subjects fear that they will be matched with an opponent playing a "strange" strategy and that only this matching would determine payoffs. Additionally, implementing three periods also allows us to test whether subjects play the same strategy for different values of the signal.

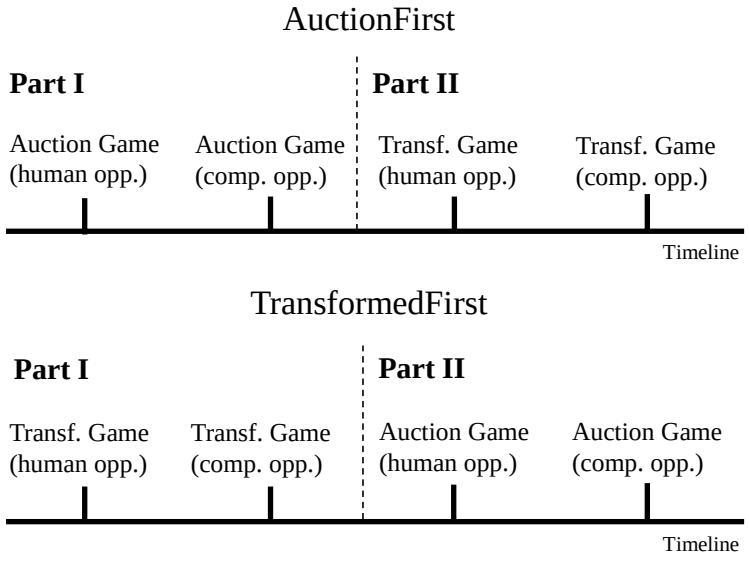


Figure 4.1: Sequence of Games in Both Treatments

the computer follows the simple strategy to just (absolutely) bid its signal. Subjects are informed about this modification only after they have finished the initial three periods. Afterwards, Part II of the experiment begins. The instructions of the transformed game are read aloud and subjects play three periods of this game with human opponents and stranger matching. Finally, subjects additionally play three periods of the transformed game with computerized opponents. In this setting, the computer follows the strategy to bid exactly the average of the lower and the higher value of the commodity ($\frac{W_1^*+W_2^*}{2}$).[12] The *TransformedFirst* treatment had a parallel design. It just changed the sequence of events such that the transformed game (with human opponents and computerized opponents) was implemented first and the auction game (with human opponents and computerized opponents) was implemented in second place. Figure 4.1 illustrates the sequence of events in both treatments.

Our experimental design allows for the following tests: First, we can test whether there is a between-subject treatment difference between the two games by comparing the initial three trading periods of Part I of both treatments. Additionally, we can check whether there is a difference when strategic uncertainty in form of human opponents is present or not. For this purpose, we can compare treatment-wise whether there is a within-subject difference of the first (human opponents) and the last (computerized opponents) three trading periods in Part I of both treatments. We did not implement the games with human and computerized opponents in reversed order. With our order, subjects should not

[12]In our experimental instructions, actually, three parts were implemented. Part II was split up into Part II and Part III. This was done in order to prevent that subjects would think about how a setting with computerized opponents would look like which might have influenced their decision already in the setting with human opponents. For ease of exposition, we do not make this distinction in our description of the treatments and the results. We just consider two parts.

learn from playing the more complex game first because no feedback is provided. Playing on the contrary the less complex game with computerized opponents first, would have potentially facilitated playing the game with human opponents. Additionally, we did not implement any belief elicitation in order not to complicate the game setting any further and in order to avoid effects from asking for beliefs on game play. Our primary interest is in subjects' actions, whether subjects are able to avoid the WC or not. Additionally, our data already allows us to reach some conclusions to what extent differences between games can be driven by differences in beliefs or not, as will become clear below. Finally, we can test whether playing the auction game first helps people understand the transformed game and vice versa. For this purpose, we can perform a within-subject analysis whether subjects behave more rationally in Part I or Part II of both treatments. Additionally, we can compare whether behavior in Part II of one treatment is similar to behavior in Part I of the other treatment because in both parts the same games are played. More details of our analysis are provided in the next section.

The experiments were conducted at the University of Mannheim in Spring 2014. Overall, six sessions with 10 to 20 subjects in each session were run. In total, 96 subjects participated. The experimental software was developed in z-Tree (Fischbacher, 2007). For recruitment, ORSEE was used (Greiner, 2004). Participants received a show-up fee of 4€. We used *Taler* as an experimental currency where each Taler was worth 0.50€. Because it is well known that subjects often loose money in CVA setting, subjects received an initial endowment of 10 Taler in Part I and II of the experiment from which losses were subtracted and to which gains were added. Even if participants made losses in both parts, they kept their initial show-up fee. Sessions lasted on average 60-75 minutes and subjects earned on average 14.40€.

4.2.3 Hypotheses

Following the idea that the problem associated with contingent reasoning on hypothetical events is at the origin of the WC as suggested by Charness and Levin (2009), this section outlines three hypotheses regarding what we expect in the data. Our major question is whether we observe a between-subject treatment difference between the auction game and the transformed game (both played in Part I of the *AuctionFirst* and the *TransformedFirst* treatment). Such a difference might first of all occur with human opponents, but it might potentially also occur with computerized opponents if the degree of rational behavior in the auction game with computerized opponents still allows for improvements. If a major obstacle in finding the equilibrium strategy or at least playing a best response to one's beliefs is the necessity to condition on the hypothetical event of winning in the WC – as suggested by Charness and Levin (2009) – we should observe that subjects in the transformed game are to a large extent able to avoid the WC:

> ***Hypothesis 1:*** A treatment difference between the auction and the transformed game (played in Part I of both treatments) should be observed. This treatment difference is expected to hold both for the setting with human and the setting with computerized opponents. When subjects do not have to condition their decisions on hypothetical events, more subjects should be able to avoid losses and hence avoid the WC.[13]

Charness and Levin (2009) suggest that subjects' problems to condition on hypothetical events is at the origin of the WC. This does, however, not preclude that even if subjects solve the conditioning problem (or the experimenter solves the problem for them), the difficulties associated with strategic interaction with human opponents (strategic uncertainty and belief formation in general) might still be a severe enough obstacle for many subjects to behave rationally. To test how important problems of strategic interaction with human opponents still are, we implement both the auction and the transformed game with computerized instead of human opponents to eliminate the problems associated with human opponents. Although problems with conditional reasoning on hypothetical events might be at the origin of the WC, we hypothesize that strategic interaction with human opponents complicates both settings.

At least two main problems with strategic interaction arise: First, subjects might fully understand the game structure but face strategic uncertainty. They do not know how others behave in this setting, how rational their opponents are. This form of strategic uncertainty might result in subjects having inconsistent beliefs about their opponent, but at least subjects best respond to these wrong beliefs. Following, however, the results of Ivanov et al. (2010), we might suspect that subjects have more fundamental problems. In a complex setting that many subjects might not fully understand, these subjects might have a more general problem to form any beliefs about their opponent at all. If subjects do not have any clue how to behave themselves in one of the two games, they might miss a starting point for forming beliefs about their opponents, from which they also could start a best-response calculation that could finally lead to the equilibrium. In order not to complicate the setting further, we did not ask for beliefs in the experiment. Hence, we cannot directly distinguish both possible explanations. Importantly, however, both explanations differ in one important prediction: With strategic uncertainty alone, subjects should not show a bidding behavior that is not a best response if we assume that subjects are uncertain about their opponents' bids.[14] If subjects have a more general problem in belief formation, however, such kind of behavior might be observed frequently.[15]

[13] In the results section, we will discuss in more detail whether the difference between the two games is really driven by a better understanding of the transformed game or whether already differences in the belief of how rationally other subjects play might be able to explain observed differences.

[14] What this means precisely, will be outlined in the results section.

[15] Obviously, the last statement assumes that subjects are able to best respond. Without this ability, behavior that is never a best response might be observed frequently even though subjects are able to form

Hypothesis 2: Comparing subjects' behavior in both games between the version with human opponents and computerized opponents within-subject (and within-treatment in Part I of both treatments) should reveal that in both games significantly more subjects are able to avoid the WC and bid closer to the equilibrium prediction if strategic uncertainty with human opponents is abolished. Importantly, with human opponents, we might frequently observe behavior that is not a best response because subjects do not only have problems to form correct beliefs but to form beliefs at all. Hence, not only a problem with contingent reasoning but also problems with belief formation lead to the WC.

Finally, the claim of Charness and Levin (2009) that problems with contingent reasoning are at the origin of the WC also suggests that we should observe a different learning pattern from Part I to Part II between the two treatments. If conditional reasoning is an obstacle for understanding the auction game, playing this game before the transformed game should not per se improve behavior in the transformed game. Subjects should not gain a better understanding of the transformed game via the auction game simply because most participants do not understand the auction game because of the problems with contingent reasoning and those subject who manage to avoid the WC in the auction game would most likely already play rationally in the transformed game if this is played first. Playing the transformed game first, however, might very well facilitate playing the auction game. By understanding the structure of the transformed game, a better understanding of the setting in which contingent reasoning on future events is necessary might arise. Hence, different patterns of learning behavior between treatments should be observed:

Hypothesis 3: In the *AuctionFirst* treatment, no learning effect should be observed. Playing the transformed game after playing the auction leads to similar results than first playing the transformed game. In the *TransformedFirst* treatment, however, a learning effect should be observed: Playing the auction game after the transformed game leads to more rational behavior than playing the auction game first.[16]

inconsistent or even consistent beliefs. Those subjects that have shown that they are able to best respond in the setting with computerized opponents should also be able to best respond in the setting with human opponents, if they only face the problem of strategic uncertainty and not a more general problem of belief formation.

[16]Our design can, however, not distinguish whether such a learning effect is driven by the fact that subjects really understand the necessity to condition on the event of winning in the WC because they played the transformed game first, or whether alternatively, subjects only learn that bidding low is a good strategy in the transformed game which they then also apply in the auction game. It is, however, noteworthy, that subjects at least do not receive any feedback about the results before the end of the experiment. Hence, they do not get any feedback on whether bidding low in the transformed game is a good strategy

4.3 Results

In this section, we first focus on Part I of the two treatments and analyze to what extent a treatment difference between the two games emerges (Section 4.3.1). Afterwards, we analyze to what extent playing one of the two games first facilitates the understanding of the other game by also looking at Part II of both treatments (Section 4.3.2).

4.3.1 Part I - Contingent Reasoning and Belief Formation

In the *AuctionFirst* treatment, subjects in Part I first play the auction game against human opponents and afterwards the same game against computerized opponents, whereas in the *TransformedFirst* treatment, subjects play the transformed game against human opponents and computerized opponents in Part I. The core evidence supporting Hypothesis 1 that subjects are to a larger extent able to avoid the WC in the transformed game is the following: In the auction game with human opponents, 61 percent of all subjects who win the game incur a loss whereas this is only the case for 32 percent of subjects in the transformed game. This finding is supported by Table 4.1. It reports the mean values for subjects' bids and subjects' average payoffs in both treatments for the two games with human and computerized opponents. Additionally, the last column of the table reports whether the distribution of bids and payoffs is different between the two treatments according to a two-sided Wilcoxon rank sum test. Finally, the last rows report whether the distribution of bids and payoffs is different within-treatment between the setting with human and computerized opponents according to a two-sided Wilcoxon signed rank test. In Table 4.1 and most of the analysis below, we average subjects' three bids for the same game to perform an analysis with independent observations. This is also justified to the extent that overall only in one out of eight games a learning effect within the three periods of a game is observed, as will be outlined in more detail in Section 4.3.2. Additionally, individual observations are independent because subjects do not receive any feedback before the end of the experiment.

We will first analyze the setting with human opponents. In this setting, average bids are significantly lower (and hence closer to the equilibrium of $b_i = -8$) in the transformed game compared to the auction game. This difference naturally also results in a significant difference in payoffs. On average, subjects lose money in the auction game whereas they win money in the transformed game. Still, even in the transformed game, subjects' payoffs are far below the average payoff which would arise from equilibrium play ($Eu_i = 2.5$).

Beyond mean values, Figure 4.2 shows histograms for subjects' bids in Part I of the *AuctionFirst* treatment and Figure 4.3 provides similar histograms for the *TransformedFirst* treatment. Considering the case of playing against human opponents first (Figure 4.2(a) and Figure 4.3(a)), the two figures reflect the aggregate finding: Subjects make lower bids more often in the transformed game than they do in the auction game. Actually, the

Table 4.1: Summary Statistics - **Part I**: Both Treatments

Means (Std. deviation)		*AuctionFirst* treatment Auction game	*TransformedFirst* treatment Transf. game	Wilcoxon rank sum (treatment difference) p-value
Human opponents	Bids	-1.80	-4.00	0.000
		(2.63)	(2.61)	
	Payoffs	-0.56	0.55	0.001
		(1.55)	(1.37)	
Comp. opponents	Bids	-3.37	-5.00	0.007
		(3.30)	(2.53)	
	Payoffs	0.17	0.81	0.004
		(1.53)	(1.56)	
Wilcoxon signed rank (within treatment)	Bids	0.000	0.020	
p-value	Payoffs	0.001	0.184	

Notes: The last column reports two-sided p-values of Wilcoxon rank sum tests that evaluates whether the distribution of bids and payoffs is different between treatments. The last rows report two-sided p-values of Wilcoxon signed rank tests that evaluate whether the distribution of bids and payoffs is different within-subject between the human and the computerized setting.

bidding behavior in the auction game to some extent gives the impression of normally distributed bids that do not reflect the equilibrium of $b_i = -8$ at all, whereas bidding behavior in the transformed game seems to at least partially reflect that the equilibrium is the lowest possible bid.

Additional light on subjects' behavior in the two different games is shed by the contingency table provided by Table 4.2. In this table, we distinguish between treatments and classify subjects in different categories according to what extent subjects follow a plausible strategy. The first category is playing the equilibrium, bidding $b_i = -8$. Subjects who bid the equilibrium seem to understand the game and follow a rational strategy. Of course, bidding -8 might not be a best response when other subjects deviate from equilibrium play.[17] Actually, even if we assume that all subjects are fully rational in their beliefs and in their behavior, we might not observe equilibrium behavior in the case that rationality is not common knowledge. If subjects for example assume that other subjects (or at least some of them) follow a naive strategy of bidding their signal (or bidding the average of the two common values), $b_j = 0$, bidding $b_i = -5.99$ would be the best response. Bearing in mind that some subjects seem to be reluctant to bid non-integer numbers, we allow our second category to be $-8 < b_i \leq -5$. Together with playing the equilibrium, we

[17] Actually, given the empirical distribution of subjects' behavior in each treatment, equilibrium play actually is not a best response. Importantly, however, best responses are very close to the equilibrium. In the auction game, bidding $b_i = -7.97$ is the best response. In the transformed game, bidding $b_i = -7.99$ is the best response. Because best responses (given the empirical distribution of subjects behavior) are so close to the equilibrium and we cannot expect subjects to tailor their behavior in cent-level deviations from the equilibrium, we will still use the equilibrium as a benchmark for rational behavior. Importantly, no subject at least on average bids the described best responses.

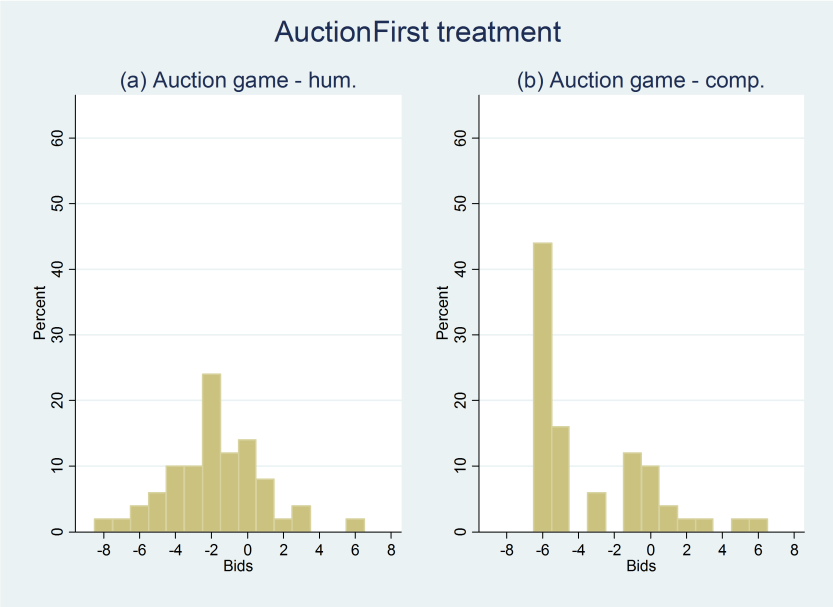


Figure 4.2: *AuctionFirst* Treatment - Bids (Part I), n = 50

still consider this kind of behavior as plausible behavior in the sense that it potentially reflects a best response to beliefs that are empirically not fully true but also not utterly implausible.

The third category is bidding $-5 < b_i \leq -3$. Such kind of behavior is only a best response if the opponent at least bids $b_j = 1$ or more. This seems to be an implausible belief to the extent that such a behavior would not even lead to a positive average payoff even if subjects do not condition in their profit calculation on the event of winning (in the auction setting). The last category is bidding $b_i > -3$. It can in principle be a weakly best

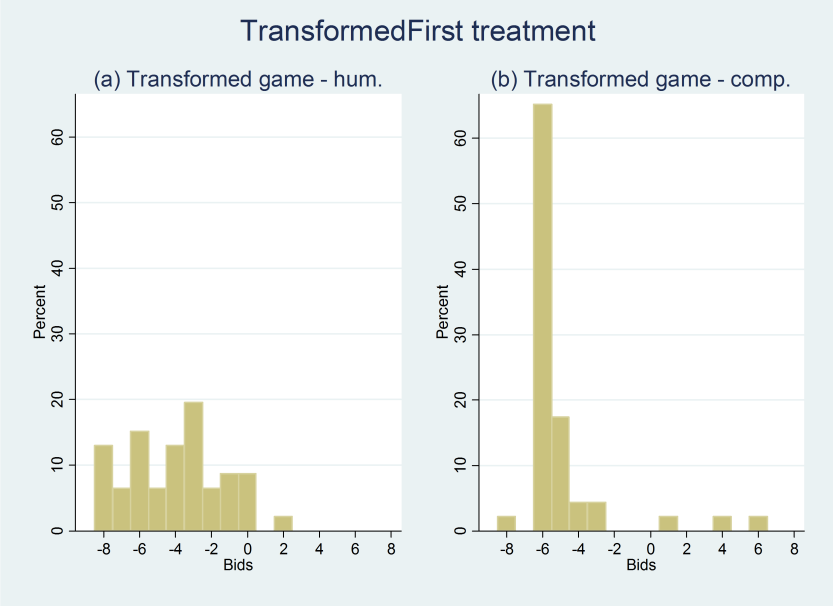


Figure 4.3: *TransformedFirst* Treatment - Bids (Part I), n = 46

Table 4.2: Contingency Table - Game with Hum. Opponents (Part I)

	Subjects' behavior				
	(1)	(2)	(3)	(4)	
Game (Treatment)	$b_i = -8$	$-8 < b_i \leq -5$	$-5 < b_i \leq -3$	$b_i > -3$	Total
Auction game	0	6	8	36	50
Transformed game	6	12	12	16	46
Total	6	18	20	52	96

response if the opponent bids $b_j > 3$ (or more precisely $b_j \geq b_i + 6$). If subjects, however, are uncertain about others' bids, or more precisely, if player i assigns at least only a small probability on the event that his opponent might not only bid $b_j > 3$ but also bid $b_j = 3$ (or even lower), bidding $b_i > -3$ can never be a best response because in this case it on average leads to losses.[18] Only bidding $b_i = -3$ (or potentially bidding even less) ensures not to make losses. Additionally, bidding $b_i = -3$ instead of $b_i > -3$ does also not lead to lower payoffs under the belief that the opponent bids $b_j > 3$ for sure. Importantly and unsurprisingly, empirically only very few subjects bid $b_j \geq 3$ as can be seen by Figure 4.2(a) and 4.3(a).

Summing it up, we think that both category (1) and (2) represents plausible behavior in the games if subjects have appropriate beliefs. Category (3) might still be a best response to some forms of implausible beliefs. Category (4) can in principle also be a best response, but it cannot be a best response if subjects are uncertain about their opponents' bids as described above. Hence, we do not consider behavior in category (3) and (4) as plausible behavior. Bearing this classification in mind, reveals that many more subjects bid plausibly (category 1 or 2) in the transformed game (39 percent) than in the auction game (12 percent), although even in this game plausible bidding behavior does not constitute the majority of all bids. In the auction game, actually, nobody bids the equilibrium. The observed differences in Table 4.2 lead to a significant difference using Fisher's exact test (two-sided: $p = 0.001$). This test is appropriate here because it compares categorical count data between independent samples. Hence, we observe a treatment difference. In the games with human opponents, subjects on average behave more rationally in the transformed game than in the auction game. Removing the obstacle of contingent reasoning on hypothetical events helps subjects to avoid the WC. Importantly, subjects really seem to understand the transformed game better than the auction game. Differences in beliefs about how other subjects behave cannot fully explain the differences

[18] If the opponent bids very high, $b_j \geq 3$, any bid $b_i \leq b_j - 6$ will be a best response as long as it ensures that player i will not win the game, as already outlined in Section 4.2.1. Naturally, any such bid will also not lead to a positive profit, given that player i's beliefs are correct. When the opponent bids $b_j = 3$ with some probability, bidding $b_i > -3$ does not ensure not to win the auction for sure anymore. And in case player i wins the auction, he will make a loss.

across games found above to the extent that in the auction game 72 percent of subjects (36 out of 50) follow a strategy that can not be a best response when we assume that subjects are uncertain about their opponents' bids (category 4) whereas this is only true for 34 percent of subjects (16 out of 46) in the transformed game. This provides first evidence that differences in the understanding of the game, not just differences in subjects' beliefs about their opponents' bids, lead to the observed treatment difference.

Next, we consider subjects' behavior in the games with computerized opponents. Figures 4.2(b) and 4.3(b) show that a large number of subjects seems to be able to find the correct solution of the games without any strategic interaction with human opponents. Additionally, the number of subjects playing the equilibrium seems again to be higher in the transformed game than in the original auction game. In line with these observations, in the auction game, 45 percent of those subjects who win the auction game face a loss whereas only 13 percent of those subjects do so in the transformed game (with computerized opponents). In Table 4.1, additional evidence supporting this result can be found: Average bids are significantly lower (and hence closer to the equilibrium $b_i = -5.99$) in the transformed game than in the auction game, whereas subjects payoffs are significantly higher.

To get further insights to subjects' behavior, Table 4.3 provides a contingency table in which we classify subjects' behavior according to four different categories. The first category, bidding $b_i \leq -6$, describes underbidding the equilibrium ($b_i = -5.99$). The second category actually is the best response against the computer's strategy ($b_j = 0$), i.e. $b_i = -5.99$. As in the case of human opponents, the third category reflects that some subjects might be reluctant to bid non-integer values. Hence, we also consider bidding $-5.99 < b_i \leq -5$ as a plausible response against the computer's strategy. Theses subject seem to understand that they have to strongly underbid the computer but they did not figure out that -5.99 is the optimal solution. Importantly, we do not consider any deviation below $b_i = -5.99$ (category 1) as plausible because even only bidding $b_i = -6$ would result in the fact that the computer would always win under both signals. This follows from the tie-breaking rule (the player with lower signal wins auction) we implemented in the auction setting. Hence, we consider both the first category, bidding $b_i \leq -6$, and the fourth category, bidding $b_i > -5$, as non-plausible behavior.

Overall, Table 4.3 shows that subjects play a plausible strategy, equilibrium (category 2) or near equilibrium play (category 3), less often in the auction game (54 percent) than in the transformed game (80 percent). Hence, even if subjects exactly know how their opponents react, conditioning on the event of winning still seems to be a problem at least for some subjects. Fisher's exact test shows that differences between the two games in Table 4.3 are statistically significant (two-sided $p = 0.012$). This provides clear evidence that subjects really understand the transformed game better than the auction game. Because the computer's strategy is known in both settings, differences in beliefs

Table 4.3: Contingency Table - Game with Comp. Opponents (Part I)

	Subjects' behavior				
	(1)	(2)	(3)	(4)	
Game (Treatment)	$b_i \leq -6$	$b_i = -5.99$	$-5.99 < b_i \leq -5$	$b_i > -5$	Total
Auction game	0	18	9	23	50
Transformed game	1	24	13	8	46
Total	1	42	22	31	96

cannot explain differences across treatment in the setting with computerized opponents.

> **Result 1:** Both with human opponents and with computerized opponents, a treatment difference between the auction game and the transformed game is observed. Hence, we find evidence in a CVA setting (with human opponents) for the claim of Charness and Levin (2009) that subjects have problems to perform contingent reasoning on hypothetical events. Importantly, the observed treatment differences cannot just be explained by differences in beliefs about what the opponents do. Subjects really seem to get a better understanding of the game in the transformed setting.

Having compared between-subjects whether a treatment difference exists or not, we now consider the question whether the necessity of belief formation in a setting with human opponents remains to be another obstacle to understand the game and avoid the WC. We again only consider Part I of both treatments in this section. In the auction game, 61 percent of the winning subjects face losses with human opponents whereas this is only the case for 45 percent with computerized opponents. In the transformed game, 32 percent of the winning subjects face losses with human opponents whereas this is only the case for 13 percent of subjects with computerized opponents. Table 4.1 supports this view: Both in the auction and the transformed game, subjects bid significantly less when facing computerized opponents. Subjects also earn higher average payoffs in both games when facing computer opponents. This difference is, however, only significant according to a two-sided signed rank test for the auction game. Importantly, we have to bear in mind that we observe differences in subjects' behavior and their payoffs even though the equilibrium requires a higher bid in the setting with computerized opponents ($b_i^{hum} = -8$ vs $b_i^{comp} = -5.99$) and even though average equilibrium profits are also lower in the computerized setting ($Eu_i = 2.5$ vs. $Eu_i = 1.495$). Or in other words, the change in equilibrium behavior biases against the observed differences that are potentially due to a better understanding when subjects are disburdened from the obstacle of belief formation.

Additional evidence that belief formation seems to be another obstacle in avoiding the WC comes from the contingency tables presented in Table 4.2 and 4.3. We first of all

only look at plausible play: as outlined before, for the setting with human opponents, we consider bidding $-8 = b_i \leq -5$ as plausible whereas we consider bidding $-5.99 = b_i \leq -5$ as plausible for the setting with computerized opponents. According to this categorization, in the auction setting, 12 percent (6 out of 50) behave plausibly with human opponents and 54 percent (27 out of 50) behave plausibly with computerized opponents. This difference is significant according to a McNemar's test (two-sided $p = 0.000$) that performs a similar test for binary categories as the Fisher's exact test does and is additionally appropriate for matched data. In the transformed game setting, 39 percent of subjects (18 out of 46) behave plausibly with human opponents but 80 percent of subjects (37 out of 46) behave plausibly with computerized opponents. This difference is again significant according to a McNemar's test (two-sided $p = 0.000$). Similar results emerge when considering equilibrium play.[19]

By construction of the treatments, subjects always play the setting with human opponents first. Still, we believe that it is very unlikely that subjects perform better in the setting with computerized opponents due to learning. Most importantly, subjects do not receive any feedback of their monetary performance until the end of the experiment. Hence, participants at least cannot learn avoiding losses by experiencing losses. Additionally, the setting with human opponents is naturally more complex than the setting with computerized opponents. Although it might be possible to gain a better understanding of a complex game by playing a simplified game before that uncovers the structure of a game setting, it is unclear how subjects should benefit in playing a simple game from playing a complex game that they do not understand. Unsurprisingly, all 6 subjects that play plausibly in the auction game with human opponents, also do so in the auction game with computerized opponents. For the transformed game, out of those 18 subjects that play plausibly without computerized opponents 17 do the same with computerized opponents. Hence, subjects who are able to understand the more complex setting with human interaction do not fail in the simpler setting with computerized opponents. But it seems most likely that they also would not have failed in the setting with computerized opponents when not playing the version with human opponents first.

Overall, the question remains *why* subjects have difficulties to avoid the WC in a setting with human opponents. Is it just a problem of strategic uncertainty in the sense that subjects understand the game structure as they do in the setting with computerized opponents but have wrong beliefs about how their opponents actually behave? Or do

[19]Considering exact equilibrium play (category 1 with human opponents and category 2 with computerized opponents), it is fairly obvious that the presence of human opponents remains a problem for playing the equilibrium strategy: In the auction game, 0 percent play the equilibrium in the game with human opponents whereas 36 percent (18 out of 50) play the equilibrium in the game with computerized opponents. This difference is significant according to a McNemar's test (two-sided $p = 0.000$). In the transformed game, 13 percent of subjects (6 out of 46) play the equilibrium in the game with human opponents whereas 52 percent (24 out of 46) play the equilibrium in the game with computerized opponents. The difference is significant according to a McNemar's test (two-sided $p = 0.000$)

subjects have more general problems in forming beliefs in the sense that in the setting with human opponents they might miss a starting point for their thoughts about the game which is provided in the setting with computerized opponents by the computer's strategy. In the *AuctionFirst* treatment, there is suggestive evidence that strategic uncertainty alone might be unable to explain the differences between settings. In this treatment, 27 out of 50 subjects are able to best respond (or nearly best respond) in the setting with computerized opponents. These subjects seem to understand the game structure well enough to best respond, at least if they are provided with a certain belief as a starting point for their thoughts about the game. Out of these 27 subjects, however, a clear majority of 17 people (63 percent) play a strategy that cannot be a best response if subjects are uncertain about their opponents' behavior in the game with human opponents (category 4). 4 people play a strategy that can only be a best response with fairly implausible beliefs (category 3 - 15 percent) and 6 people play a plausible strategy (categories 1 and 2 - 22 percent). This suggests that a majority of subjects who are in principle able to best respond face problems of belief formation that go beyond strategic uncertainty.

In the *TransformedFirst* treatment, the number of participants who are able to best respond in the setting with computerized opponents is higher: 37 subjects. Out of these 37 subjects, 11 subjects follow a strategy that can not be a best response in the setting with human opponents if we assume that subjects are uncertain about their opponents' behavior (category 4 - 30 percent). 9 subjects follow a strategy that can only be a best response with fairly implausible beliefs (category 3 - 24 percent). 11 and 6 subjects follow a plausible strategy (category 1 and 2 - 30 and 16 percent). Hence, although the results are not as clear as in the *AuctionFirst* treatment, still the behavior of a majority of subjects could only be rationalized in terms of strategy uncertainty when we assume that subjects have fairly implausible beliefs. It seems that in a setting with human opponents subjects seem to miss a belief starting-point for their thoughts about the game structure. Hence, they are unable to avoid strategies that can not be a best responses when subjects are uncertain about their opponents' bids (or that can be a best response only with fairly implausible beliefs) although they are in principle able to calculate the best response in the setting in which a specific belief is given. Overall:

> **Result 2:** Belief formation in a setting with human opponents is an obstacle for avoiding the WC both in the auction and in the transformed game. Although Charness and Levin (2009) are right that contingent reasoning on hypothetical events prevents many subjects from behaving rationally, the comparison between the situation with human and computerized opponents shows that strategic interaction with human opponents provides another obstacle to avoid the WC. Importantly, this is also true for the transformed game in which the problem of conditional reasoning has already been removed. Additionally, the problem of strategic uncertainty alone seems not to be able to rationalize

the differences between the two settings. A general problem of belief formation seems to be present in the data.

A remaining question is whether the conditioning problem or the belief formation problem is a greater obstacle for avoiding the WC and playing rationally. To analyze this question, we compare the auction game with computerized opponents (no belief formation problem but a conditioning problem) with the transformed game with human opponents (belief formation problem but no conditioning problem). In the auction game (with computerized opponents) 45 percent of subjects who win the game face losses whereas this is only the case for 32 percent of those subjects in the transformed game (with human opponents). Table 4.1 supports this finding: In the transformed game with human opponents, subjects bid on average slightly lower than in the auction game with computerized opponents (-4.00 vs. -3.37), which also results in slightly higher average payoffs in this game (0.55 vs. 0.17). Importantly, however, the differences in bidding behavior and payoffs are not significant according to a Wilcoxon rank sum test (Bids - two-sided $p = 0.771$; Payoffs - two-sided $p = 0.322$). Moreover, the comparison is misleading because the equilibrium bid (equilibrium payoff) is always higher (lower) in the game with computerized opponents, as outlined before. Hence, analyzing whether the percentage of plausible behavior changes between games is a much more reliable measure. This analysis, however, does also not reveal a difference between the two games (Fisher's exact test using bidding $-8 \leq b_i < -5$ and $-5.99 \leq b_i < -5$ as a classification for plausible behavior - two-sided $p = 0.158$). Only a similar Fisher's exact test using the exact equilibria as a classification device does lead to significant difference (two-sided $p = 0.035$): 36 percent (18 out of 50) of subjects play the equilibrium when facing only the problem of contingent reasoning whereas only 13 percent (6 out of 46) play the equilibrium when facing only the belief formation problem. Overall, however, we have the following result:

Result 3: In a CVA setting, the problem of contingent reasoning on hypothetical events and the problem of belief formation in an environment with human opponents seem to provoke the WC in a similar magnitude. If there is a difference in magnitude at all, however, the problem of belief formation seems to lead to slightly more severe problems (Fisher's exact test - exact equilibria). In any case, it is true that the problem of conditioning on the event of winning is not the only obstacle to avoid the WC in CVA setting.

4.3.2 Part I and Part II - Learning

In this section, we will analyze whether playing a particular game first facilitates playing the other game. Before we do this, some general remarks about learning in our treatments. As outlined before, subjects play all games in our setting for three periods each. Because

subjects do not receive any feedback about their payoffs until the end of the experiment, they cannot learn in these three periods to avoid losses by experiencing losses. One way of still learning in these three periods would e.g. be that subjects take their last period bid as a belief of how others bid in the current period. We test whether subjects indeed learn during the three periods they play each game by testing whether the distribution of bids for the first and the last of the three periods is different. Over both treatments, this test can be done for eight games. Only in one game, a statistically significant result (Wilcoxon signed rank test - two sided: $p = 0.001$) is observed: the transformed game played first in the *TransformedFirst* treatment. Subjects on average bid -3.17 in the first period and -4.68 in the third period.

This result seems plausible to the extent that in both settings with computerized opponents we would expect that due to the simplicity of the setting subjects either understand the problem immediately or they fail to understand it also after playing three periods. Additionally, in the auction setting with human opponents, the conditioning problem might prohibit learning during the three periods. Only the transformed game seems complex and at the same time simple enough so that subjects can improve their performance when playing the game for three periods. This, however, only seems to be true when playing this game as the first game in the *TransformedFirst* treatment. When playing the game only in Part II in the *AuctionFirst* treatment, subjects do not show any significant improvement.[20] For completeness, Figures C.1 and C.2 (in Appendix C.2) provide individual bids for all 12 periods of the experiment for all subjects of the two treatments. These figures support the evidence presented so far that subjects only improve their behavior in the transformed game when this game is played in Part I of the

[20]Statistical results presented so far were based on averaging bids for three periods. Or central results, however, remain fairly robust when looking at individual periods: When we compare bidding behavior and payoffs between the auction game and the transformed game (both with human opponents) and this time base this comparison only on the first period, the treatment difference is slightly less significant: The Wilcoxon rank sum test for the bidding behavior remains significant (two sided: $p = 0.018$) and payoff differences remain at least marginally significant (two sided: $p = 0.050$). A Fisher's exact test using bidding below -5 as a classification devices is significant (two sided: $p = 0.011$), whereas such a test using the exact equilibrium as a classification device is not significant (two sided: $p = 0.109$). Because subjects in the transformed game still improve their behavior over the three periods, our comparison with the auction game based on the first period leads to less pronounced results. This, however, seems not to question our results. Naturally, comparing the auction and the transformed game based on the third period would of course lead to a more pronounced difference than doing comparison with mean values. When we compare bidding behavior and payoffs in the transformed game between the setting with human opponents and computerized opponents we may wonder whether we still observe a difference when comparing behavior in the last period of the game with human opponents and the first period with computerized opponents. Actually, differences in the bidding and payoff distributions vanish (Wilcoxon sign rank tests - bids: two-sided $p = 0.376$; payoffs: two-sided $p = 0.766$). We have, however, to bear in mind that the equilibrium predictions between those settings changes which biases against a possible finding that behavior between settings is different. Hence, the much more reliable measure of comparison is the McNemar's test (that incorporates the change in equilibrium) either based on exact equilibria (two sided: $p = 0.002$) or on behavior classified as plausible (two sided: $p = 0.004$). These tests remain highly significant. Hence, overall, we conclude that using single periods instead of mean values does not qualitatively change the results gained so far.

Table 4.4: Summary Statistics - **Part II**: Both Treatments

Mean (Std. deviation)		*AuctionFirst* treatment Transf. game	*TransformedFirst* treatment Auction game
Human opponents	Bids	-3.66	-3.79
		(4.05)	(2.88)
	Payoffs	0.05	0.29
		(2.29)	(1.90)
Comp. opponents	Bids	-3.04	-4.48
		(3.74)	(2.66)
	Payoffs	-0.16	0.68
		(2.09)	(1.53)

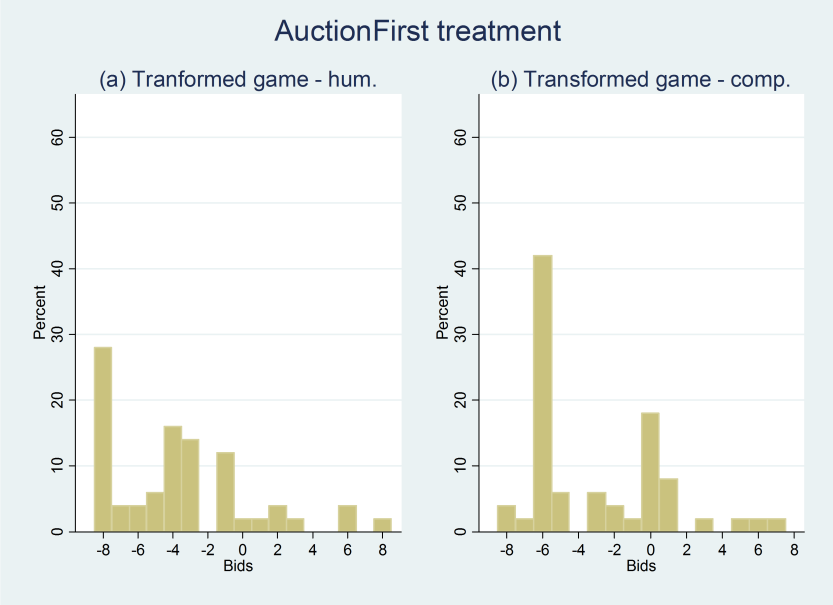


Figure 4.4: *AuctionFirst* Treatment - Bids (Part II), n = 50

experiment.

But does playing one game first facilitates playing the other game? Table 4.4 provides the mean values for subjects' bids and payoffs for Part II of both treatments.[21] In the *AuctionFirst* treatment, the transformed game was played in the second part, both with human opponents and computerized opponents. In the *TransformedFirst* treatment, the auction game was played in the second part, again both with human opponents and computerized opponents. Figure 4.4 and 4.5 additionally show histograms of subjects' bids in the *AuctionFirst* and the *TransformedFirst* treatment for Part II of both treatments.

For the *TransformedFirst* treatment, we hypothesized that we might observe a learning effect. We will look at the setting with human opponents first: When the auction game is

[21] We omit the non-parametric tests shown in Table 4.1 because for the analysis we would like to perform in this section mainly tests comparing results in Part I with results in Part II are necessary.

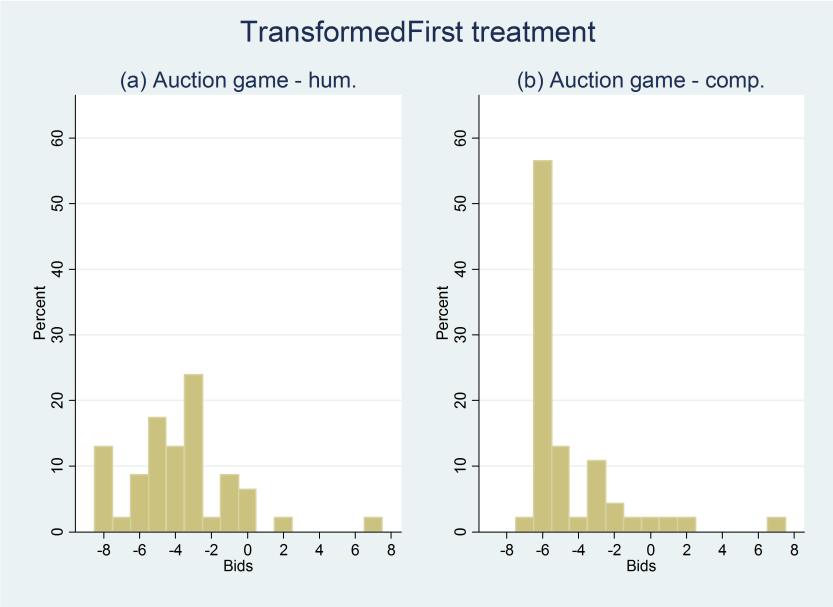


Figure 4.5: *TransformedFirst* Treatment - Bids (Part II), n = 46

played after the transformed game (*TransformedFirst* treatment), only 28 percent of those subjects who win the game face losses, whereas 61 percent of those subjects face losses when the auction is played first (*AuctionFirst* treatment). In line with this observation, bids in auction game are lower in *TransformedFirst* treatment (Figure 4.5(a)) than in the *AuctionFirst* treatment(Figue 4.2(a)), whereas payoffs are higher (Mean values - bids: -3.79 vs. -1.80; payoffs $+0.29$ vs. -0.56).[22] Hence, there is clear evidence that playing the transformed game in the *TransformedFirst* treatment before the auction game helps subjects to avoid the WC in the auction game. Because of learning, we also do not observe the treatment effect between the two games within-subject in the *TransformedFirst* treatment: Bids and payoffs are roughly the same between the transformed and the auction game in this treatment (Mean values - bids: -4.00 vs. -3.79; payoffs: 0.55 vs. 0.29).[23]

Do we also observe this learning effect in the *TransformedFirst* treatment for the setting with computerized opponents? When the auction game is played after both transformed games (*TransformedFirst* treatment), only 22 percent of those subjects who win the game face losses, whereas 45 percent of those subjects face losses when the auction game is played in Part I (of the *AuctionFirst* treatment). In line with this observation, bids in auction game (with computerized opponents) are lower in *TransformedFirst* treatment (Figure 4.5(b)) than in the *AuctionFirst* treatment (Figure 4.2(b)), whereas payoffs are higher

[22]Wilcoxon rank sum test - bids: two-sided $p = 0.000$; payoffs: two-sided $p = 0.002$. Fisher's exact test classifying bids equal or below -5 as rational - two-sided p-value $= 0.025$. Again, a Fisher's exact test using the exact equilibrium as a classification threshold do not lead to different results (two-sided: $p = 0.022$).

[23]Wilcoxon signed rank test - bids: two-sided $p = 0.814$; payoffs: two-sided $p = 0.833$. Additionally, a McNemar's test (two-sided: $p = 0.6072$) classifying subjects as rational that bid equal or below -5 reveals no significant difference. Again, a McNemar's exact test using the exact equilibrium as a classification threshold do not lead to different results (two-sided $p = 1.000$).

(Mean values - bids: -4.48 vs. -3.37; payoffs $+0.68$ vs. $+0.17$). Hence, it again looks like that subjects behave slightly more rationally when they play the transformed game first compared with the situation when this is not the case. Statistical support, however, provides only partial support for this this impression.[24] Additionally, unlike in the case of human opponents, the learning effect seems not to be strong enough to totally prevent a treatment effect also within-subject.[25] Hence, there is some evidence for a learning effect also in the *TransformedFirst* treatment, but this learning effect seems to be weaker than in the setting with human opponents. In the *TransformedFirst* treatment, the auction game with computerized opponents was played as the last game. Potentially, exhaustion or confusion because of all the different games played before might have been highest at the end of the experiment, diminishing the learning effect. At least, subjects behave less rational than expected in the very last game of the *TransformedFirst* treatment.

For the *AuctionFirst* treatment, we hypothesized above that subjects should not benefit from playing the auction game first in playing the transformed game second. We will first analyze the setting with human opponents: When the transformed game is played after the auction game (*AuctionFirst* treatment), 47 percent of those subjects who win the game face losses, whereas 32 percent of those subjects face losses when the transformed game is played first (*TransformedFirst* treatment). Additionally, bids in the transformed game are even slightly higher in *AuctionFirst* treatment (Figure 4.4(a)) than in the *TransformedFirst* treatment(Figure 4.3(a)), whereas payoffs are lower (mean values - bids: -3.66 vs. -4.00; payoffs $+0.05$ vs. $+0.55$). Differences, however, are small and not statistical significant.[26][27]

[24] Wilcoxon rank sum test - bids: two sided $p = 0.076$; payoffs: two sided $p = 0.054$. But: Fisher's exact test (bidding $-5.99 < b_i \leq -5$ as a classification device) - two sided: $p = 0.301$. A Fisher's exact test using the exact equilibrium as a classification threshold does not lead to different results (two-sided $p = 0.412$).

[25] Again, the statistical analysis is fairly inconclusive. A Wilcoxon signed rank test just reveals no significant difference (bids: two-sided $p = 0.101$; payoffs: two-sided $p = 0.371$) within-subject between the transformed and the auction game (with computerized opponents), but a McNemar's test classifying subjects as rational that bid equal or below -5 reveals such a difference with marginal significance (two-sided p-value $= 0.065$), whereas a McNemar's test using the exact equilibrium as a classification device does again not reveal this difference (two-sided p-value $= 1.000$).

[26] Wilcoxon rank sum test: Bids - two-sided $p = 0.848$; payoffs - two-sided $p = 0.293$. Additionally, a Fisher's exact test using bids smaller or equal -5 as a classification criterion for plausible behavior supports this finding (two-sided $p = 0.834$). Surprisingly, even a Fisher's exact test using the exact equilibrium does not reveal any difference between subjects behavior (two-sided $p = 0.293$) although Figure 4.4(a) suggests a high level of equilibrium play. Importantly, however, the "-8"-bin in this figure also captures lots of bids that are very close to the equilibrium but that are not exactly -8.

[27] Difference additionally remain statistically insignificant (with the exception of payoffs) when only comparing bids and payoffs for the last of the three periods (and not mean values for all three periods) and hence controlling for the learning which takes place in the transformed game when played first in the *TransformedFirst* treatment: Wilcoxon rank sum test: bids (last period) - two-sided $p = 0.306$; payoffs (last period) - two-sided $p = 0.061$. Additionally, a Fisher's exact test using (last period) bids smaller or equal -5 as a classification criterion for plausible behavior supports this finding (two-sided $p = 0.209$). Also a Fisher's exact test using the exact equilibrium does not reveal any difference between subjects behavior (two-sided $p = 0.478$). Hence, subjects in the *AuctionFirst* treatment do not perform better in the transformed game than subjects in the *TransformedFirst* treatment but, importantly, they also do not perform worse, which might have been an indication that the randomization of subjects did not work.

In any case, subjects do not seem to learn how to avoid the WC in the transformed game from playing the auction game first. Because subjects do not learn in the *AuctionFirst* treatment, we also observe the treatment effect between the two games within-subject in this treatment: Bids are higher in the auction game compared to the transformed game, whereas payoffs are lower (mean values - bids: -1.80 vs. -3.66; payoffs: -0.56 vs. $+0.05$)[28]

How does the behavior in the games with computerized opponents evolve in the *AuctionFirst* treatment? When the transformed game is played after both auction games (*AuctionFirst* treatment), 43 percent of those subjects who win the game face losses, whereas only 13 percent of those subjects face losses in the transformed game in Part I of the *TransformedFirst* treatment. In line with this observation, bids in transformed game (with computerized opponents) are higher in *AuctionFirst* treatment (Figure 4.4(b)) than in the *TransformedFirst* treatment (Figure 4.3(b)), whereas payoffs are lower (Mean values - bids: -3.04 vs. -5.00; payoffs -0.16 vs. $+0.81$)[29] Hence, in the setting with computerized opponents, we do not only not observe a learning effect, but subjects in the *AuctionFirst* treatment even perform slightly worse than in the *TransformedFirst* treatment. For this reason, we also do not observe the treatment effect between the two games within-subject in the *AuctionFirst* treatment: Bids and payoffs are fairly similar in the auction game compared to the transformed game (mean values - bids: -3.37 vs. -3.04; payoffs: $+0.17$ vs. -0.16).[30]

As in the *TransformedFirst* treatment, learning behavior seems to be slightly different between the setting with human opponents and computerized opponents also in the *AuctionFirst* treatment. Our - admittedly - speculative explanation why this is the case is

---

Additionally, in the *AuctionFirst* treatment, one might argue that there is a different kind of learning effect in the sense that subjects do not perform better in the transformed game than subjects in the *TransformedFirst* treatment, but at least these subjects do not have to learn over the three periods of the game (as in the *TransformedFirst* treatment) because the auction game was played before. Importantly, however, differences between treatments in the transformed game are also not significant when comparing first round behavior which potentially speaks against this different kind of learning: Wilcoxon rank sum test: bids (first period) - two-sided $p = 0.274$; payoffs (first period) - two-sided $p = 0.652$. Additionally, a Fisher's exact test using (first period) bids smaller or equal -5 as a classification criterion for plausible behavior supports this finding (two-sided $p = 0.302$). Also a Fisher's exact test using the exact equilibrium does not reveal any difference between subjects behavior (two-sided $p = 0.130$).

[28]Wilcoxon signed rank tests: bids - two-sided $p = 0.000$; payoffs - two-sided $p = 0.003$. This result is also supported by a McNemar's test (two-sided $p = 0.002$) using bids below -5 as a classification criterion for plausible behavior. A McNemar's test using the exact equilibrium as classification criterion for rational behavior leads to a similar result (two sided $p = 0.001$).

[29]Wilcoxon rank sum test: bids - two-sided $p = 0.033$; payoffs - two-sided $p = 0.007$. Fisher's exact test using bids $-5.99 < b_i \leq -5$ as a classification for rational behavior - two-sided $p = 0.001$. This time, a Fisher's exact test using the exact equilibrium as a classification for rational behavior leads to different result (two-sided $p = 0.150$). Although this result would be more in line with our Hypothesis, we believe that bidding $-5.99 < b_i \leq -5$ is the more appropriate criterion for rational behavior as outlined above.

[30]Wilcoxon signed rank test: bids - two-sided $p = 0.980$; payoffs - two-sided $p = 0.205$. McNemar's test using bidding $-5.99 < b_i \leq -5$ as a criterion for plausible behavior - two-sided $p = 0.549$. A McNemar's test using the exact equilibrium as a classification criterion for rational play leads to a similar result (two sided: $p = 1.000$).

the following: As in the *AuctionFirst* treatment, the transformed game with computerized opponents was played as the last game. First of all, subjects might already be slightly exhausted at this point of the experiment. In addition, when solving this game they might at least consider two other games as a reference: the transformed game with human opponents and the auction game with computerized opponents. Considering both games might have lead to some confusion of at least some subjects, leading e.g. to the very high frequency of zero bids in the transformed game with computerized opponents (imitating the computer's strategy - Figure 4.4(b)). At least, as in the *TransformedFirst* treatment, we also observe in *Auction treatment* that subjects behave less rational than expected in the very last game of the experiment. Overall:

> ***Result 4:*** In the setting with human opponents, we observe a learning effect as hypothesized: Playing the transformed game first facilitates playing the auction game, whereas the reverse is not true. With computerized opponents, a similar but weaker learning effect is observed in the *TransformedFirst* treatment. Overall, however, rationality levels in the last game of both treatments are lower than expected. Exhaustion or increased confusion might be responsible for this result.

4.4 Conclusion

In this chapter, we analyze whether the claim of Charness and Levin (2009) that subjects have problems to perform contingent reasoning on hypothetical events is at the origin of the WC seems to be valid in a CVA setting in which human opponents are present. Charness and Levin (2009) argue that the conditional reasoning problem is important by observing the WC even in a setting in which subjects do not face human opponents but face computerized opponents that follow a known strategy. We, however, directly manipulated whether subjects face a conditioning problem in a setting in which human opponents are present. For this purpose, we use a simplified auction game to construct a number choosing game (framed as an auction) in which subject do not have to condition on the future event of winning the game which they have to do in the auction setting.

We obtain two central results: First, without the conditioning problem, subjects are to a larger extent able to avoid the WC and this result holds in an environment with human opponents which has not be shown in the literature so far. Second, even without the conditioning problem, many subjects fail to avoid the WC in the setting with human opponents. We observe that the necessity to form beliefs in a setting with human opponents seems to be another obstacle for subjects in avoiding the WC. Our data suggests that both problems, contingent reasoning and belief formation, have a negative impact of similar magnitude on the rationality of subjects' behavior in our games. Charness and Levin

(2009) are right that contingent reasoning on future events is a serious problem in the CVA setting. Nonetheless, the problem of belief formation in this setting also cannot be underestimated.

Finally, it seems worthwhile to test to what extent our findings for the WC and the CVA setting also hold in similar environments in which contingent reasoning and belief formation about other human players are necessary. Esponda and Vespa (2014) have come to fairly similar conclusion than Charness and Levin (2009) in an experimental setting. Hence, analyzing the impact of contingent reasoning and belief formation in a voting setting by construction a transformed voting game that disburdens subjects at least partially from contingent reasoning seems a worthwhile area of future research.

Bibliography

Abbink, Kaus, Jordi Brandts, Benedikt Herrmann, and Henrik Orzen, "Intergroup Conflict and Intra-Group Punishment in an Experimental Contest Game," *American Economic Review*, 2010, *100(1)*, 420–447.

Abbink, Klaus and Abdolkarim Sadrieh, "The Pleasure of Being Nasty," *Economics Letters*, 2009, *105*, 306–308.

_ **and Benedikt Herrmann**, "The Moral Costs of Nastiness," *Economic Inquiry*, 2011, *49(2)*, 631–633.

Agnell, Jonas and Per Lundborg, "Theories of Pay and Unemployment: Survey Evidence from Swedish Manufactoring Firms," *Scandinavian Journal of Economics*, 1995, *97*, 295–307.

Akerlof, George, "The Market for Lemons: Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 1970, *84*, 488–500.

_ , "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 1982, *97*, 543–569.

Akerlof, Gerorge and Janet Yellen, "The Fair Wage-Effort Hypothesis and Unemployment," *Quarterly Journal of Economics*, 1990, *105*, 255–283.

Algan, Yann and Pierre Cahuc, "Inherited Trust and Growth," *American Economic Review*, 2010, *100(5)*, 2060–2092.

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 1989, *97(6)*, 1447–1458.

_ , "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal*, 1990, *100 (June)*, 464–477.

Anscombe, Gertrude E.A., "Modern Moral Philosophy," *Philosophy*, 1958, *124(1)*, 1–19.

Argyle, Michael, "Causes and Correlates of Happiness," in Daniel Kahneman and Norbert Schwartz, eds., *Well-Being - The Foundations of Hedonic Psychology*, Russel Sage Foundation, 1999.

Aristotle, *The Nichomachean Ethics (translated by J.E.C. Welldon)*, Prometheus Books, 1987.

Arrow, Kenneth J., "Gifts and Exchanges," *Philosophy & Public Affairs*, 1972, *1(4)*, 343–362.

Avery, Christopher and John H. Kagel, "Second-Price Auctions with Asymmetric Payoffs: An Experimental Investigation," *Journal of Economics & Management Strategy*, 1997, *6(3)*, 573–603.

Azariadis, Costas, "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, 1975, *83*, 1183–1202.

Baily, Martin, "Wages and Employment under Uncertain Demand," *Review of Economic Studies*, 1974, *41*, 37–50.

Banfield, Edward, *The Moral Basis of a Backward Society*, New York: Free Press, 1958.

Bartling, Björn and Klaus M. Schmidt, "Reference Points, Social Norms, and Fairness in Contract Renegotiations," *Journal of the European Economic Association*, 2014. forthcoming.

Batson, C., J.L. Dyck, J.R. Brandt, J.G. Batson, A.L. Powell, M.R. McMaster, and C. Griffitt, "Five Studies Testing Two New Egoistic Alternatives to the Empathy-Altruism Hypothesis," *Journal of Personality and Social Psychology*, 1988, *55(1)*, 52–77.

Bazerman, Max H. and Wiliam F. Samuelson, "I Won the Auction But Don't Want the Prize," *Journal of Conflict Resolution*, 1983, *27*, 618–634.

Becchetti, Leonardo and Giacomo Antoni, "The Sources of Happiness: Evidence from the Investment Game," *Journal of Economic Psychology*, 2010, *31(4)*, 498–509.

_ , Luisa Corrado, and Pierluigi Conzo, "Sociability, Altruism and Subjective Well-Being," *CEIS Tor Vergata Research Papper Series*, 2013, *11(5)*.

Becker, P., *Psychologie der seelischen Gesundheit. Band 1: Theorien, Modelle, Diagnostik*, Göttingen: Hogrefe, 1982.

Bewley, Truman, *Why Wages Don't Fall During a Recession*, Cambridge, MA: Havard University Press, 1999.

Blanchard, Olivier J. and Jordi Gali, "Real Wage Rigidities and the New Keynesian Model," *Journal of Money, Credit, and Banking*, 2007, *39(1)*, 35–69.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann, "A Within-Subject Analysis of Other-Regarding Preferences," *Games and Economic Behavior*, 2011, *72* (2), 321 – 338.

Blinder, Alan and Don Choi, "A Shred of Evidence on Theories of Wage Stickiness," *Quarterly Journal of Economics*, 1990, *105*, 1003–1015.

Bohnet, Iris, Nava Ashraf, and Nikita Piankov, "Decomposing Trust and Trustworthiness," *Experimental Economics*, 2006, *9.3*, 193–208.

Bolton, Gary and Axel Ockenfels, "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 2000, *90(1)*, 166–193.

_ , Jordi Brands, and Axel Ockenfels, "Fair Procedures: Evidence from Games Involving Lotteries," *The Economic Journal*, 2005, *115*, 1054–1076.

Bosman, Ronald and Frans van Winden, "Emotional Hazard in a Power-to-Take Experiment," *The Economic Journal*, 2002, *112* (476), 147–169.

Bradburn, Norman, *The Structure of Psychological Well-Being*, Chicago: Aldine Publishing Company, 1969.

Brandts, Jordi and Gary Charness, "Do Labor Market Conditions Affect Gift Exchange? Some Experimental Evidence," *Economic Journal*, 2004, *114*, 684–708.

_ , _ , and Matthew Ellman, "Let's Talk: How Communication Affects Contract Design," *Mimeo*, 2012.

Brown, Kirk and Richard Ryan, "The Benefits of Being Present: Mindfulness and Its Role in Psychological Well-Being," *Journal of Personality and Social Psychology*, 2003, *84(4)*, 822–848.

Bruni, Luigino and Robert Sugden, "Reclaiming Virtue Ethics for Economics," *Journal of Economic Perspectives*, 2013, *27(4)*, 141–164.

Bull, Clive, "The Existence of Self-Enforcing Implict Contracts," *Quarterly Journal of Economics*, 1987, *102*, 142–159.

Burchardi, Konrad B. and Stefan Penczynski, "Out Of Your Mind: Eliciting Individual Reasoning in One Shot Games," *Games and Economic Behavior*, 2014, *84*, 39–57.

Burda, Michael, Werner Güth, Georg Kirchsteiger, and Harald Uhlig, "Employment Duration and Resistance to Wage Reductions: Experimental Evidence," *Homo Oeconomicus*, 2005, *22(2)*, 169–183.

Burks, Stephen, Jeffrey Carpenter, and Eric Verhoogen, "Playing Both Roles in the Trust Game," *Journal of Economic Behavior & Organization*, 2003, *51*, 195–216.

_ , Jon Anderson, Jeffrey Carpenter, Lorenz Götte, Karsten Maurer, Danielle Nosenzo, Ruth Potter, Kim Rocha, and Aldo Rustichini, "Self Selection Does Not Increase Other-Regarding Preferences among Adult Laboratory Subjects, but Student Subjects May Be More Self-Regarding than Adults," *IZA Discussion Paper No. 5389*, 2010.

Camerer, Colin F., "The Promise and Sucess of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List," *Mimeo*, 2011.

Campbell, Carl and Kunal Kamlani, "The Reasons for Wage Rigidity: Evidence from a Survey of Firms," *Quarterly Journal of Economics*, 1997, *112*, 759–789.

Charness, Gary and Brit Grosskopf, "Relative payoffs and Happiness: An Experimental Study," *Journal of Economic Behavior & Organization*, 2001, *45* (3), 301 – 328.

_ and Dan Levin, "The Origin of the Winner's Curse: A Laboratory Study," *American Economic Journal, Microeconomics*, 2009, *1(1)*, 207–236.

_ and Martin Dufwenberg, "Promises and Partnership," *Econometrica*, 2006, *74*, 1579–1601.

_ and Matthew Rabin, "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 2002, *117(3)*, 817–869.

_ , Ramón Cobo-Reyes, Natalia Jiménez, Juan A. Lacomba, and Francisco Lagos, "The Hidden Advantag of Delegation: Pareto Improvements in a Gift Exchange Game," *American Economic Review*, 2012, *102(5)*, 2358–2379.

Chen, Daniel L. and John J. Horton, "The Wages of Pay Cuts: Evidence from a Field Experiment," *Mimeo*, 2009.

Christoffel, Kai and Tobias Linzert, "The Role of Real Wage Rigidity and Labor Market Frictions for Inflation Persistence," *Journal of Money, Credit, and Banking*, 2010, *42(7)*, 1435–1446.

Crawford, Vincent P. and Nagore Iriberri, "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 2007, *75(6)*, 1721–1770.

de Quervain, Dominique, Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Urlich Schnyder, Alfred Buck, and Ernst Fehr, "The Neural Basis of Altruistic Punishment," *Science*, 2004, *305*, 1254–1258.

Diener, Ed and Martin Seligman, "Beyond Money: Toward an Economy of Well-Being," *Psychological Science in the Public Interest*, 2004, *5*, 1–31.

_ , Robert Emmons, Randy Larsen, and Sharon Griffin, "The Satisfaction with Life Scale," *Journal of Personality Assessment*, 1985, *49(1)*, 71–75.

Easterlin, Richard, "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," in P. David and M. Reder, eds., *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz*, New York and London: Academic Press, 1974.

Eliaz, Kfir and Ran Spiegler, "Reference Dependence and Labor-Market Fluctuations," *NBER Macroeconomics Annual*, 2013, *28.* forthcoming.

Esponda, Ignacio and Emanuel Vespa, "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal, Microeconomics*, 2014. forthcoming.

Eyster, Erik and Matthew Rabin, "Cursed Equilibrium," *Econometrica*, 2005, *73(5)*, 1623–1672.

Falk, Armin, Ernst Fehr, and Christian Zehnder, "Fairness Perceptions and Reservation Wages - The Behavioral Effects of Minimum Wage Laws," *Quarterly Journal of Economics*, 2006, *121*, 1347–1381.

_ , _ , and Urs Fischbacher, "On The Nature of Fair Behavior," *Economic Inquiry*, 2003, *41(1)*, 20–26.

Fehr, Ernst, Alexander Klein, and Klaus M. Schmidt, "Fairness and Contract Design," *Econometrica*, 2007, *75*, 121–154.

_ and Armin Falk, "Wage Rigidity in a Competitive Incomplete Contract Market," *Journal of Political Economy*, 1999, *107(1)*, 106–134.

_ and Klaus Schmidt, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 1999, *114(3)*, 817–868.

_ **and Lorenz Goette**, "Robustness and Real Consequences of Nominal Wage Rigidity," *Journal of Monetary Economics*, 2005, *52*, 779–804.

_ **and Urs Fischbacher**, "Why Social Preferences Matter - The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives," *Economic Journal*, 2002, *112*, C1–C33.

_ **and** _ , "Third-Party Puishment and Social Norms," *Evolution and Human Behavior*, 2004, *25*, 63–87.

_ **, Georg Kirchsteiger, and Arno Riedl**, "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics*, 1993, *108*, 437–459.

_ **, Lorenz Goette, and Christian Zehnder**, "A Behavioral Account of the Labor Market: The Role of Fairness Constraints," *Annual Review of Economics*, 2009, *1*, 355–384.

_ **, Oliver Hart, and Christian Zehnder**, "Contracts as Reference Points - Experimental Evidence," *American Economic Review*, 2011, *101(2)*, 493–525.

_ **,** _ **, and** _ , "How Do Informal Agreements and Revisions Shape Contractual Reference Points?," *Journal of the European Economic Association*, 2014. forthcoming.

_ **, Urs Fischbacher, and Michael Kosfeld**, "Neuroeconomic Foundations of Trust and Social Preferences," *American Economic Review*, 2005, *95(3)*, 346–351.

Fischbacher, Urs, "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 2007, *10(2)*, 171–178.

_ **, Simon Gächter, and Ernst Fehr**, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters*, 2001, *71*, 397–404.

Fleming, Piers and Daniel John Zizzo, "Social Desirability, Approval and Public Good Contribution," *Personality and Individual Differences*, 2011, *51(3)*, 258–262.

Foot, Philippa, *Virtues and Vices and Other Essays in Moral Philosophy*, Oxford: Blackwell, 1978.

Frederick, Shane, "Cognitive Reflection and Decision Making," *Journal of Economc Perspectives*, 2005, *19(4)*, 25–42.

Frey, Bruno S. and Alois Stutzer, "What Can Economists Learn from Happiness Research," *Journal of Economic Literature*, 2002, *XL*, 402–435.

Fukuyama, Francis, *Trust: The Social Virtues and The Creatin of Prosperity*, Free Press, 1995.

Gächter, Simon and Arno Riedl, "Moral Property Rights in Bargaining with Infeasible Claims," *Management Science*, 2005, *51(2)*, 249–263.

_ **, Elke Renner, and Martin Sefton**, "The Long-Run Benefits of Punishment," *Science*, 2008, *322 (5907)*, 1510.

Gerhards, Leonie and Matthias Heinz, "In Good Times and Bad - Reciprocal Behavior at the Workplace in Times of Economic Crises," *Mimeo*, 2012.

Gintis, Herbert, "The Nature of Labor Market Exchange and the Theory of Capitalist Production," *Review of Radical Political Economics*, 1976, *8*, 36–54.

_ , "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology*, 2000, *206*, 169–179.

_ **, Samuel Bowles, Robert Boyd, and Ernst Fehr**, "Explaining Altruistic Behavior in Humans," *Evolution and Human Behavior*, 2003, *24*, 153–172.

Goeree, Jakob K. and Theo Offerman, "Efficiency in Auctions and Public Information in Common Values: An Experimental Study," *American Economic Review*, 2002, *92(3)*, 347–361.

Gordon, Donald, "A Neo-Classical Theory of Keynesian Unemployment," *Economic Inquiry*, 1974, *12*, 431–459.

Gossling, Samuel D., Peter J. Rentfrow, and William B. Swann Jr., "A Very Brief Measure of the Big-Five personality domains," *Journal of Research in Personality*, 2003, *37*, 504–528.

Graham, Liam and Andrew Oswald, "Hedonic Capital, Adaption and Resilience," *Journal of Economic Behavior & Organization*, 2010, *76*, 372–384.

Greiner, Ben, "An Online Recruitment System for Economic Experiements," in Kurt Kremer and Volker Macho, eds., *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, Ges. für Wiss. Datenverarbeitung, 2004. 79-93.

Grosskopf, Brit, Yoella Bereby-Meyer, and Max Bazerman, "On the Robustness of the Winner's Curse Phenomenon," *Theory and Decision*, 2007, *63*, 389–418.

Guven, Cahit, "Are Happier People Better Citizens?," *Kyklos*, 2011, *64(2)*, 178–192.

Hall, Rober E., "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 2005, *95(1)*, 50–64.

Hannan, Lynn, "The Combined Effect of Wages and Firm Profit on Employee Effort," *The Accounting Review*, 2005, *80(1)*, 167–188.

Hart, Oliver and John Moore, "Contracts as Reference Points," *Quarterly Journal of Economics*, 2008, *123(1)*, 1–48.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter, "Antisocial Punishment Across Societies," *Science*, 2008, *319*, 1362–1367.

Herweg, Fabian and Klaus M. Schmidt, "A Theory of Inefficient Ex-Post Renegotiation," *Mimeo*, 2012.

Hirsch and Peterson, "Extraversion, Neuroticism, and the Prisoner's Dilemma," *Personality and Individual Differences*, 2009, *46*, 254–256.

Hoff, Karla, Mayuresh Kshetramade, and Ernst Fehr, "Caste and Punishment: The Legacy of Caste Culture in Norm Enforcement," *The Economic Journal*, 2011, *121*, F449–F475.

Ifcher, John and Homa Zarghamee, "Happiness and Time Preferences: The Effect of Positive Affect in a Random-Assigment Experiment," *American Economic Review*, 2011, *101(7)*, 3109–29.

Ivanov, Asen, Dan Levin, and Muriel Niederle, "Can Relaxation of Beliefs Rationalize the Winner's Curse?: An Experimental Study," *Econometrica*, 2010, *78(4)*, 1435–1452.

James, Harvey, "Is the Just Man a Happy Man? An Empirical Study of the Relationship between Ethics and Subjective Well-Being," *Kyklos*, 2011, *64(2)*, 193–212.

Jonckheere, Aimable Robert, "A Distribution-Free k-Sample Test against Ordered Alternatives," *Biometrika*, 1954, *41(1/2)*, 133–145.

Jones, Alvin and Rick Crandall, "Validation of a Short Index of Self-Actualization," *Personality and Social Psychology Bulletin*, 1986, *12(1)*, 63–73.

Kagel, John H. and Dan Levin, "The Winner's Curse and Public Informatin in Common Value Auctions," *American Economic Review*, 1986, *76(5)*, 894–920.

Kahneman, Daniel, Ed Diener, and Norbert Schwartz, eds, *Well-being: The Foundations of Hedonic Psychology*, New York: Russel Sage, 1999.

_ , Jack L. Knetsch, and Richard Thaler, "Fairness as Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, 1986, *76(4)*, 728–741.

Karlan, Dean and John A. List, "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment," *American Economic Review*, 2007, *97(5)*, 1774–1793.

Kesebir, Pelin and Ed Diener, "A Virtuous Cycle: The Relationship Between Happiness and Virtue," *Mimeo*, 2013.

Keyes, Corey, "Social Well-Being," *Social Psychology Quarterly*, 1998, *61(2)*, 121–140.

Keyes, Corey L.M., Carol D. Ryff, and Dov Shmotkin, "Optimizing Well-Being: The Empirical Encounter of Two Traditions," *Journal of Personality and Social Psychology*, 2002, *82(6)*, 1007–1022.

Kirchsteiger, Gerorg, Luca Rigotti, and Aldo Rustichini, "Your Morals Might Be Your Moods," *Journal of Economic Behavior & Organization*, 2006, *Vol. 59*, 155–172.

Knack, Stephen and Philip Keefer, "Does Social Capital Have an Economic Payoff? A Cross-Country Comparison," *Quaterly Journal of Economics*, 1997, *112*, 1251–1288.

Knoch, Darina, Alvaro Pascual-Leone, Kaspar Meyer, Valerie Treyer, and Ernst Fehr, "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex," *Science*, 2006, *314*, 829–832.

Koch, Christian, "Tax Compliance, the Treat of National Insolvency and Well-Being," *Dissertation Proposal*, 2011.

Kocher, Martin and Sebastian Strasser, "The Fair Employment Hypothesis: Reciprocity in Unstable Environments," *Mimeo*, 2011.

Konow, James, "Mixed Feelings: Theories of and Evidence on Giving," *Journal of Public Economics*, 2010, *94* (3-4), 279 – 297.

_ **and Joseph Earley**, "The Hedonistic Paradox: Is Homo Economicus Happier?," *Journal of Public Economics*, 2008, *92* (1-2), 1 – 33.

Kosfeld, Michael, Markus Heinrichs, Paul Zak, Urs Fischbacher, and Ernst Fehr, "Oxytocin Increases Trust in Humans," *Nature*, 2005, *435*, 673–676.

Krishna, Vijay, *Auction Theory*, sec. ed., Academic Press, 2009.

Krohne, H. W., B. Egloff, C.-W. Kohlmann, and A. Tausch, "Untersuchung mit einer deutschen Form der Positiven and Negative Affect Schedule (PANAS)," *Diagnostica*, 1996, *42*, 139–156.

Köszegi, Botond and Mathew Rabin, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 2006, *121(4)*, 1133–1165.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe, "Do Wage Cuts Damage Work Morale? Evidence from a Field Experiment," *Journal of the European Economic Association*, 2013, *11 (4)*, 853–870.

Kurzban, Robert and Daniel Houser, "Individual Differences in Cooperation in a Circular Public Goods Game," *European Journal of Personality*, 2001, *15*, 37–52.

Layard, Richard, *Happiness: Lessons from a New Science*, New York: The Penguin Press, 2005.

Lück, Helmut and Ernst Timaeus, "Skalen zur Messung manifester Angst (MAS) und sozialer Wünschbarkeit (SDS-E und SDS-CM)," *Diagnostica*, 1969, *15*, 134–141.

Ledyard, John O, "Public Goods: A Survey of Experimental Economic Research," in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics*, Princeton University Press, 1995, chapter 2, pp. 111–194.

Levin, Dan and Philipp J. Reiss, "Could we overcome the Winner's Curse by (behavioral) auction desing?," *Mimeo*, 2012.

Levitt, Steven D. and John A. List, "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," *Journal of Economic Perspectives*, 2007, *21(2)*, 153–174.

Lind, Barry and Charles Plott, "Thw Winner's Curse: Experiments with Buyers and Sellers," *American Economic Review*, 1991, *81(1)*, 335–346.

Lindbeck, Assar and Dennis Snower, *The Insider-Outsider Theory of Employment and Unemployment*, Cambridge, Mass.: MIT Press, 1988.

Lyubomirsky, Sonja, Kennon Sheldon, and David Schkade, "Pursuing Happiness: The Architecture of Sustainable Change," *Review of General Psychology*, 2005, *9(2)*, 111–131.

_ , Rene Dickerhoof, Julia Boehm, and Kennon Sheldon, "Becoming Happier Takes Both a Will and a Proper Way: An Experimental Longitudinal Intervention to Boost Well-Being," *American Psychological Association*, 2011, *11(2)*, 391–402.

Mortensen, Dale, "Job Search and Labor Market Analysis," in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*, Vol. vol. 2, Amsterdam: North-Holland, 1986.

Mortensen, Dale T. and Christopher A. Pissarides, "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 1994, *61*, 397–415.

Muck, Peter M., Benedikt Hell, and Samuel D. Gossling, "Construct Validation of a Short Five-Factor Model Instrument," *European Journal of Psychological Assessment*, 2007, *23(3)*, 166–175.

Newbery, David M. and Joseph E. Stiglitz, "Wage Rigidity, Implicit Contracts, Unemployment and Economic Efficiency," *Economic Journal*, 1987, *97*, 416–430.

Oechssler, Jörg, Andreas Roider, and Patrick Schmitz, "Cognitive Abilities and Behavioral Biases," *Journal of Economic Behavior & Organization*, 2009, *72*, 147–152.

Riedl, Arno and Jean-Robert Tyran, "Tax Liability Side Equivalence in Gift-Exchange Labor Markets," *Journal of Public Economics*, 2005, *89*, 2369–2382.

Riggi, Marianna, "Nominal and Real Wage Rigidities in New Keynesian Models: A Critical Survey," *Journal of Economic Surveys*, 2010, *24(3)*, 539–572.

Ryan, Richard and Edward Deci, "On Happiness and Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being," *Annual Review of Psychology*, 2001, *52*, 141–166.

Ryff, Carol, "Happiness Is Everything, or Is It? Exploarations on the Meaning of Psychological Well-Being," *Journal of Personality and Social Psychology*, 1989, *Vol. 57, No. 5*, 1069–1081.

_ **and Burton Singer**, "Know Thyself and Become What You Are: A Eudaimonic Approach to Psychological Well-Being," *Journal of Happiness Studies*, 2008, *9(1)*, 13–39.

Sanfey, Alan, James Rilling, Jessica Aronson, Leigh Nystrom, and Jonathan Cohen, "The Neural Basis of Economic Decision-Making in the Ultimatum Game," *Science*, 2003, *300*, 1755–1758.

Schumacher, Jörg, Antje Klaiberg, and Elmar Brähler, eds, *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden*, Göttingen: Hogrefe, 2003.

Shapiro, Carl and Joseph Stiglitz, "Equilibrium Unemployment as a Worker Discipline Devise," *American Economic Review*, 1984, *74*, 433–444.

Sheldon, Kennon and Sonja Lyubomirsky, "Achieving Sustainable Gains in Happiness: Change Your Actions, Not Your Circumstances," *Journal of Happiness Studies*, 2006, *7(1)*, 55–86.

_ **and Tim Kassner**, "Coherence and Congruence: Two Aspects of Personal Integration," *Journal of Personality and Social Psychology*, 1995, *68(3)*, 531–545.

Shimer, Robert, "The Cyclical Behavior of Equilibrium Unemploymet and Vacancies," *American Economic Review*, 2005, *95*, 25–49.

Tabellini, Guido, "Institutions and Culture," *Journal of the European Economic Association*, 2008, *6(2-3)*, 255–294.

van 't Wout, Mascha, René Kahn, Alan Sanfey, and André Aleman, "Affective State and Decision-Making in the Ultimatum Game," *Experimental Brain Research*, 2006, *169*, 564–568.

Volk, Stefan, Christian Thöni, and Winfried Ruigrok, "Personality, Personal Values and Coopertion Preferences in Rublic Goods Games: A Longitudinal Study," *Personality and Individual Differences*, 2011, *50(6)*, 810–815.

Watson, David, Anna Clark, and Auke Tellegen, "Development and Validation of Brief Mesarues of Positive and Negative Affect: The PANAS Scales," *Journal of Personality and Social Psychology*, 1988, *54(6)*, 1063–1070.

Weiss, Andrew, "Job Queues and Layoffs in Labor Markets with Flexible Wages," *Journal of Political Economy*, 1980, *88*, 526–538.

Zheng, Liu and Louis Phaneuf, "Technology and Labor Market Dynamics: Some Evidence and Theory," *Journal of Monetary Economics*, 2007, *54(8)*, 2534–2553.

Appendix A

Appendix to Chapter 2

In this appendix, I first provide three additional Tables (Appendix A.1). Afterwards, four pieces of additional analysis are presented (Appendix A.2). Then, all well-being questions used in the experiment are outlined (Appendix A.3). Finally, I provide a translated version of the experimental instructions (Appendix A.4).

A.1 Additional Table

Table A.1a: Summary of Ordered Logit Results - **KE's Controls**

Dependent variable/	**Less virtuous**		**More virtuous**		**Most virtuous**	
Well-being measure		p-value		p-value		p-value
Subjective well-bing						
Long-run SWB	1.08	0.16	0.74	0.27	1.07	0.15
Highest/Lowest Happiness	0.06	0.93	0.00	0.99	0.00	0.99
short-run SWB	-0.82	0.29	-0.50	0.47	-0.86	0.26
Change in short-run SWB	0.39	0.61	0.81	0.25	1.50	**0.05****
Satisfaction With Life (SWL)	-0.96	0.19	-0.42	0.51	-0.86	0.22
Psychological well-being						
PWB	1.57	**0.04****	1.63	**0.02****	2.46	**0.00****

Notes: */** indicate significance at 10/5% level; The category least virtuous subject is not included as a regression dummy; All *p*-values below 10 % are in bold type, n=102.

Table A.1b: Summary of Ordered Logit Results - **Full Set of Demographic Controls**

Dependent variable/	**Less virtuous**		**More virtuous**		**Most virtuous**	
Well-being measure		p-value		p-value		p-value
Subjective well-bing						
Long-run SWB	0.96	0.22	0.96	0.19	1.41	0.07
Highest/Lowest Happiness	0.28	0.73	0.36	0.64	0.56	0.49
short-run SWB	-0.97	0.22	-0.36	0.63	-0.78	0.32
Change in short-run SWB	0.54	0.51	0.78	0.29	1.54	**0.05***
Satisfaction With Life (SWL)	-1.03	0.17	-0.38	0.57	-0.85	0.25
Psychological well-being						
PWB	1.73	**0.03****	1.97	**0.01****	2.76	**0.00****

Notes: */** indicate significance at 10/5% level; The category least virtuous subject is not included as a regression dummy; All *p*-values below 10 % are in bold type, n=102.

Table A.1c: Summary of Ordered Logit Results - **Full Set of Demographic Controls and Big Five Inventory**

Dependent variable/ Well-being measure	**Less virtuous**	p-value	**More virtuous**	p-value	**Most virtuous**	p-value
Subjective well-bing						
Long-run SWB	0.85	0.29	1.00	0.20	1.28	0.15
Highest/Lowest Happiness	0.52	0.52	0.28	0.72	0.76	0.38
short-run SWB	-0.84	0.30	-0.50	0.51	-0.94	0.25
Change in short-run SWB	0.69	0.41	0.81	0.28	1.44	**0.08***
Satisfaction With Life (SWL)	-1.83	**0.02****	-0.99	0.19	-2.41	**0.01****
Psychological well-being						
PWB	1.82	**0.02****	2.00	**0.01****	2.98	**0.00****

Notes: */** indicate significance at 10/5% level; The category least virtuous subject is not included as a regression dummy; All *p*-values below 10 % are in bold type, n=102.

A.2 Additional Analysis

In the following, I provide four pieces of additional analysis. First, I provide a more detailed check of instruments. Second, I provide an *aggregated* regression analysis validating the summary statistics presented in Section 2.4.1. Third, I provide a complementary *non-aggregate* regression analysis. Forth, I provide a more detailed analysis of the virtuousness hypothesis.

A.2.1 Check of Instruments and Well-Being Measures

This subsection provides a more detailed analysis of whether behavior in the games is in line with expectations and whether well-being measures show reasonable correlations with each other (and the Marlowe-Crowne scale). Besides the JOD-G, all my other games replicate the conventional results of the literature. In my DG, 61% of subjects gave whereas 39% were nongivers (KE's result: 40% nongivers).[1] In my SPD, 64% of the subjects trusted and 72% of the subjects are classified as conditional cooperators, 5% as unconditional cooperators, and 23% as free-riders (Burks et al. 2010: 55% trust, 59% conditional cooperators, 4% unconditional cooperators, 35% free-riders). In my analysis, I merge conditional and unconditional cooperators because of the low frequency of unconditional cooperators. The reader may bear in mind that the overwhelming fraction of these cooperators are conditional cooperators. In my Mini-UG, 53% of subjects rejected the unfair offer (Falk et al. 2003: 44% rejection). In my SPD-P, about 50% punished free-riders whereas only 3% punished cooperators. That is, I observe a high level of altruistic punishment but a negligible level of anti-social punishment which is in line with the very low level of spiteful preference in the JOY-G. In the DG-P, about 25% punished dictators who did not share equally. Results are roughly in line with Fehr and Fischbacher (2004): Especially a high-cost punishment technology (in the DG-P) is related to lower

[1] 39% nongivers are indeed a fairly high result for a non-double-blind setting. However, playing both roles may have an effect of *reduced responsibility* for the partner, as has been shown by Burks et al. (2003).

rates of punishment compared to a low-cost technology (in the SPD-P).

Are well-being measures also in line with results in the literature? Table A.2a shows the Spearman-rank correlation coefficients (and p-values) for subjective and psychological well-being measures.[2] All correlations have the expected sign and out of 91 correlations, 72 are significant at the 5% level. In comparison to KE (p. 16), results are slightly less significant. Taking, however, into account that I only use a single blind instead of a double blind procedure, I conclude that well-being measures show reasonable correlation results and are in line with expectations from the literature. Importantly, the measure with the highest amount of non-significant correlations is Bradburn's Negative Affect (NA): For this measure, only seven out of thirteen corrleations are significant.

Table A.2a summarizes the Spearman-rank correlation of SWB and PWB measures with the Marlowe-Crowne social desirability scale (MC), the two material well-being (MWB) measures, and cognitive ability (CRT). Nine out of fourteen SWB and PWB measures significantly correlate with the MC scale (at the 5% level). Compared to KE's result, correlations (at least for PWB) are slightly higher. Observed correlations with MC seem reasonable and indicate that social desirability plays a more important role in a single blind than in a double blind setting. In my regression analysis (Appendix A.2.2), I will, however, controll for the MC scale. Unlike in KE's result, the two material well-being (MWB) variables also significantly correlate with some well-being measures. This is, however, only the case for three out of fourteen well-being measures for each MWB measure. Additionally, for no well-being measure both material well-being measures significantly correlate with the same measure. Cognitive ability as measured by the Cognitive Reflection Test (CRT), does not show significant correlations with well-being measures. This result is in line with the happiness literature. So overall, well-being measures are in line with results from the literature.

[2]Now Happiness (NH) and Mood Index (MI) refer in this table to the first measurement of these variables in the overall well-being questionnaire at the beginning of the experiment. For this table, I also state the often reported Affect Balance Scale (ABS). It is gained by subtracting NA from PA.

Table A.2a: Spearman Correlation Matrix for SWB and PWB Measures (Correlations with Coefficients and p-values)

	OH	PA	NA	ABS	PAS	NAS	NH1	MI1	HH	LH	SWL	PWBI	SAI
Overall Happiness (OH)													
Bradburn's positive affect (PA)	0.07												
	0.457												
Bradburn's negative affect (NA)	-0.20	-0.09											
	0.05	0.39											
Affect Balance Scale (ABS)	0.16	0.69	-0.76										
	0.115	**0.000**	**0.000**										
Positive Affect Schedule (PAS)	0.39	0.43	-0.24	0.44									
	0.000	**0.000**	**0.015**	**0.000**									
Negative Affect Schedule (NAS)	-0.14	-0.23	0.42	-0.45	-0.21								
	0.172	**0.021**	**0.000**	**0.000**	**0.037**								
Now Happiness (NH)	0.35	0.26	-0.12	0.25	0.25	-0.24							
	0.000	**0.007**	0.239	**0.011**	**0.012**	**0.016**							
Mood Index (MI)	0.38	0.29	-0.26	0.37	0.38	-0.31	0.64						
	0.000	**0.003**	**0.008**	**0.000**	**0.000**	**0.002**	**0.000**						
Highest Happiness (HH)	0.40	0.51	-0.09	0.39	0.40	-0.13	0.56	0.45					
	0.000	**0.000**	0.359	**0.000**	**0.000**	0.204	**0.000**	**0.000**					
Lowest Happiness (LH)	0.28	0.15	-0.42	0.42	0.25	-0.45	0.34	0.51	0.32				
	0.005	0.140	**0.000**	**0.000**	**0.011**	**0.000**	**0.000**	**0.000**	**0.001**				
Satisfaction with Life (SWL)	0.51	0.13	-0.16	0.20	0.29	-0.25	0.33	0.27	0.33	0.19			
	0.000	0.187	0.114	**0.040**	**0.004**	**0.010**	**0.001**	**0.005**	**0.001**	0.059			
Psych. Well-Being Index (PWBI)	0.45	0.28	-0.09	0.24	0.46	-0.31	0.22	0.37	0.28	0.22	0.56		
	0.000	**0.004**	0.369	**0.014**	**0.000**	**0.002**	**0.025**	**0.000**	**0.004**	**0.029**	**0.000**		
Self-Actualization Index (SAI)	0.16	0.16	-0.30	0.32	0.32	-0.43	0.10	0.11	0.11	0.24	0.32	0.44	
	0.114	0.112	**0.002**	**0.001**	**0.001**	**0.000**	0.327	0.293	0.264	**0.014**	**0.001**	**0.000**	
Social Well-Being (SoWB)	0.32	0.20	-0.12	0.22	0.31	-0.30	0.15	0.23	0.22	0.23	0.25	0.46	0.48
	0.001	**0.048**	0.249	**0.028**	**0.002**	**0.002**	0.127	**0.021**	**0.023**	**0.022**	**0.010**	**0.000**	**0.000**

Notes: p-values below the 5% level are in bold type, n=102. Compare KE (Table 1).

Table A.2b: Spearman Correlation Matrix for SWB and PWB with MC Social Desirability Scale, Material Well-Being, Cognitive Ability (Coefficients and p-values)

	OH	PA	NA	ABS	PAS	NAS	NH1	MI1	HH	LH	SWL	PWBI	SAI	SoWB
Marlowe-Crowne scale (MC)	0.13	0.12	-0.26	0.27	0.33	-0.36	0.17	0.21	0.15	0.28	0.15	0.33	0.42	0.24
	0.187	0.217	**0.009**	**0.006**	**0.001**	**0.000**	0.090	**0.032**	0.142	**0.004**	0.142	**0.001**	**0.000**	**0.014**
Monthly expenditures	0.27	-0.10	-0.10	0.02	0.24	0.04	0.02	0.13	0.05	0.05	0.19	0.17	0.22	0.06
	0.006	0.342	0.324	0.872	**0.016**	0.711	0.846	0.190	0.587	0.643	0.062	0.079	**0.029**	0.572
Parents' income	0.17	0.14	0.06	0.04	0.15	-0.07	0.14	0.10	0.24	-0.10	0.38	0.20	-0.02	-0.02
	0.085	0.172	0.555	0.675	0.135	0.482	0.166	0.338	**0.015**	0.333	**0.000**	**0.049**	0.873	0.868
Cognitive Ability (CRT)	0.11	0.19	-0.01	0.14	0.03	-0.01	-0.08	-0.04	0.00	-0.12	0.06	0.09	0.02	-0.04
	0.291	0.059	0.957	0.166	0.736	0.937	0.421	0.721	0.979	0.218	0.573	0.379	0.808	0.716

Notes: p-values below the 5% level are in bold type, n=102. Compare KE (Table 2).

A.2.2 Aggregate Regression Analysis

The summary statistics presented in Section 2.4.1 do not control for demographic variables and personality traits. In order to validate these results, I follow KE (p. 22) and run multivariate regressions, especially ordered logit. As KE, I will skip Bradburn's Negative Affect (NA) for these regressions because the NA correlations might be an artifact, as suggested by KE.[3] Because well-being measures are highly correlated with each other, KE run regressions for each well-being measure separately to avoid the problem of multicollinearity. In these regressions, one well-being measure is treated as the dependent variable that is explained by a Giver dummy and controls. I deviate from this analysis in two important ways. First, I include additional demographic control variables and the Big Five Inventory which are missing in KE's analysis. Second, I aggregate well-being measures. This is done for two purposes: First, it facilitates the exposition of results for all six different settings a lot. Second, overall in my data the relationship between well-being and virtuous behavior is weaker than in KE's analysis. Hence, I am not so much interested in which single variable correlates with virtuous behavior but whether the different (aggregated) well-being concepts show such a correlation or not, when controls are included. Additionally, Appendix A.2.3 provides a non-aggregate regression analysis, supporting the aggregate results.

The aggregation of well-being measures is done in the following way: SWB consists of two components: an affective part and a cognitive-evaluative part. I follow Keyes et al. (2002) in calculating an aggregate *long-run SWB* and an aggregated *short-run SWB* score. For conceptual reasons, I give equal weight to both components.[4] Additionally, an aggregated *change in short-run SWB* is calculated and Highest and Lowest Happiness (HH and LH) are summed up to *Highest/Lowest Happiness*. For PWB, all three measures are aggregated giving equal weight to each individual measure.[5]

Table A.3a-A.3c provide a summary of ordered logit results. The depended variables are listed in the first column. The other columns report the coefficients (and p-values) for different dummies of virtuous behavior. Because both well-being measures and different types of virtuous behavior are highly correlated among each other, for each dependent

[3] As outlined earlier (Appendix A.2.1), NA has the least amount of significant correlations with other well-being variables and its performance in my (and KE's) analysis is rather poor. KE suggest that the NA correlations could be an artifact because NA does not seem to measure what it should measure (see KE's footnote 25 for details).

[4] For the long-run measure, I use Overall Happiness (OH), the Positive Affect Schedule (PAS) and Negative Affect Schedule (reversely scored). I do not use Bradburn's Positive and Negative Affect (PA/NA) for my aggregate score because of the problems of NA discussed earlier. For short-run happiness, Now Happiness (NH) and the Mood Index (MI) are used.

[5] Keyes et al. (2002) just sum up different well-being variables. Because my well-being measures have different scales, I first standardize those measure by subtracting the mean and dividing by the standard error before aggregating different scales. For PWB, a factor analysis of PWB measures results in one factor with very similar factor loadings for different PWB measures, supporting the approach to give equal weights to each measure.

Table A.3a: Summary of Ordered Logit Results Coefficients for All Games with Similar Controls as in KE

Dependent variable/ Well-being measure	Giver Dummy	p-value	Trustor Dummy	p-value	Cooperator Dummy	p-value	Rejector Dummy	p-value	Punisher Dummy *(SPD-P)*	p-value	Punisher Dummy *(DG-P)*	p-value
Subjective well-bing												
Long-run SWB	-0.07	0.86	1.13	**0.02****	0.34	0.40	-0.41	0.25	-0.09	0.80	0.07	0.87
Highest/Lowest Happiness	-0.24	0.52	0.20	0.60	-0.30	0.49	-0.04	0.91	-0.42	0.25	-0.19	0.65
Short-run SWB	0.12	0.75	-0.52	0.24	-0.08	0.85	-0.24	0.49	-0.28	0.43	-0.12	0.78
Change in short-run SWB	0.22	0.55	0.26	0.52	1.12	**0.01****						
Satisfaction With Life (SWL)	0.02	0.95	-0.22	0.61	-0.06	0.89	-0.28	0.44	0.11	0.76	0.32	0.45
Psychological well-being												
PWB	0.71	**0.06***	0.89	**0.05****	1.25	**0.00****	-0.35	0.33	0.36	0.32	0.92	**0.03****

Notes: */** indicate significance at 10/5% level; *p*-values below 10% are in bold type, n=102.

Table A.3b: Summary of Ordered Logit Results Coefficients for All Games with Demographic Controls

Dependent variable/ Well-being measure	Giver Dummy	p-value	Trustor Dummy	p-value	Cooperator Dummy	p-value	Rejector Dummy	p-value	Punisher Dummy *(SPD-P)*	p-value	Punisher Dummy *(DG-P)*	p-value
Subjective well-bing												
long-run SWB	-0.03	0.94	1.23	**0.01****	0.19	0.67	-0.34	0.35	-0.15	0.69	-0.19	0.66
Highest/Lowest Happiness	-0.09	0.82	0.70	0.13	-0.30	0.52	-0.05	0.89	-0.45	0.25	-0.19	0.66
short-run SWB	0.10	0.79	-0.42	0.36	0.01	0.98	-0.30	0.41	-0.37	0.32	-0.14	0.76
Change in short-run SWB	0.26	0.52	0.21	0.62	1.13	**0.02****						
Satisfaction With Life (SWL)	0.13	0.74	-0.27	0.55	-0.08	0.87	-0.33	0.38	0.07	0.85	0.21	0.63
Psychological well-being												
PWB	0.69	**0.08***	0.91	**0.04****	1.25	**0.01****	-0.39	0.29	0.30	0.42	0.85	**0.05****

Notes: */** indicate significance at 10/5% level; *p*-values below 10% are in bold type, n=102.

Table A.3c: Summary of Ordered Logit Results Coefficients for All Games with Demographic Controls and Big Five Inventory

Dependent variable/ Well-being measure	Giver Dummy	p-value	Trustor Dummy	p-value	Cooperator Dummy	p-value	Rejector Dummy	p-value	Punisher Dummy *(SPD-P)*	p-value	Punisher Dummy *(DG-P)*	p-value
Subjective well-bing												
long-run SWB	-0.33	0.43	1.33	**0.01****	0.17	0.71	-0.45	0.22	-0.40	0.29	-0.30	0.50
Highest/Lowest Happiness	-0.12	0.78	0.86	**0.08***	-0.53	0.27	0.16	0.67	-0.50	0.20	0.12	0.79
short-run SWB	0.06	0.89	-0.55	0.24	0.01	0.98	-0.25	0.50	-0.40	0.28	0.04	0.93
Change in short-run SWB	0.19	0.66	0.29	0.52	1.13	**0.02****						
Satisfaction With Life (SWL)	-0.02	0.97	-1.02	**0.04****	-0.19	0.70	-0.17	0.65	-0.23	0.55	-0.08	0.85
Psychological well-being												
PWB	0.64	0.13	1.04	**0.03****	1.51	**0.00****	-0.17	0.64	0.18	0.63	0.90	**0.04****

Notes: */** indicate significance at 10/5% level; *p*-values below 10% are in bold type, n=102.

variable and for each dummy of virtuous behavior a separate regression is run. Hence, every coefficient represents a single regression. In Table A.3a, these regressions include only basic controls replicating KE's analysis: MWB1, MWB2, the MC scale, a dummy for the sequence of games, a dummy for subjects' performance in the CRT, and subjects' level of confidence for truthfully answering well-being questions expressed in the concluding questionnaire. Table A.3b presents the same analysis with an extended set of demographic controls for which the literature has shown that they potentially have an impact on well-being (age, age squared, sex, religion, health dummy, partnership dummy, ethnicity dummy, field-of-study dummy). Table A.3c uses the same controls as in Table A.3b and additionally includes the five dimensions of the Big Five Inventory (extraversion, agreeableness, emotional stability, conscientiousness, openness to experience). Regressions for the trust decision also include risk preferences and people's expectation about what others return in the SPD. By including the last variable, I try to focus on what might explain trust beyond monetary self-interest.

A drawback of the aggregate well-being measures is that the regression coefficients (or potentially odds-ratio not shown in the table) cannot be easily interpreted as for single well-being measures because they refer to an aggregated and weighted measure. Looking however at the p-values, we see that the regression results confirm the main insights gained by the summary statistics in Section 2.4.1. For giving, trusting and cooperating, only (aggregate) PWB is robustly correlated with virtuous behavior across games and different controls. (Aggregate) long-run SWB is only correlated with trusting, reflecting the weakness of this relationship and that the Trustor dummy also showed the best correlations with long-run SWB in Table 2.2a. So regression results confirm that the PWB-giver relationship extents to trusting and cooperating in the SPD. The main deviation between the DG and the SPD, however, is also confirmed: the change in short-run SWB is significantly correlated (at least) with cooperating.

The regressions for the punishment settings confirm results gained so far: Only the relationship between (aggregate) PWB and punishing in the DG-P (TP-punishment with high costs) is significant. Punishment in the Mini-UG, however, is not positively related to PWB. In contrast to Table 2.2a-2.2b, long-run SWB is not related to rejecting in the Mini-UG. Table 2.2b, however, already shows that there is only a negative relationship between rejecting and the affective component of long-run SWB. Because the aggregate measure of long-run SWB does include both components of SWB, I do not see this relationship in Table A.3a-A.3c. The non-aggregate regression analysis of Appendix A.2.3, however, seems to support this relationship.

Finally, comparing Table A.3a-A.3b to Table A.3c, we see that including the Big Five Inventory (and hence controlling for personality characteristics beyond demographic controls) changes results for (aggregate) PWB slightly: The Giver dummy becomes insignificant. Nonetheless, the overall message seems to be that even if we control for personality

characteristics (aggregate) PWB is significant in three out of four expected situations. Additionally, including the Big Five Inventory might again give rise to multicollinearity potentially biasing the estimates. So overall, virtuous behavior in different settings seems to be robustly related to PWB.

Overall, the aggregate regression analysis suggests that KE's main finding of a strong PWB-virtuousness relationship (and a weak SWB-virtuousness relationship) extents from giving in the DG to trusting and cooperating in the SPD. Additionally, presented evidence suggests that there is a crucial distinction between SP- and TP-punishment.

A.2.3 Non-Aggregate Regression Analysis

In Appendix A.2.2, an aggregate regression analysis of the relationship between well-being and virtuous behavior has been provided. In this section, I try to extend this analysis by a similar non-aggregate regression analysis. The major message of this analysis is the following: Significance of the aggregate PWB measure in the aggregate regression analysis (Table A.3a-A.3c) is mainly driven by the SAI and SoWB, not by the PWBI. For SWB, some measures occasionally become significant in the non-aggregate analysis for the DG and the SPD (especially when only a limited number of controls is used), but there seems to be no measure that is robustly significant across this two games reflecting results of the aggregate analysis. Only for my punishment settings, measures of positive and negative affect seem to be related across games, as will be discussed in more detail below.

Following KE, Table A.4a-A.4f present a summary of regression results for each game. The regressions performed are very similar to those done in Table A.3a-A.3c. The only difference is that now non-aggregate well-being measures (as in Table 2.2a-2.2b) are used. In each regression, one well-being measure is the dependent variable that is explained by measures of virtuous behavior (giver dummy, trustor dummy, cooperator dummy and punisher dummy) and controls. For each well-being measure, six different ordered logit regressions are run (specification 1-6) and each time only the coefficient of the virtuous behavior dummy and the dummy's p-value are reported. For all Tables a-f, specification 1) *All* includes the same set of controls as Table A.3a and hence replicates KE's analysis most closely. Specification 2) *Conf6*, restricts the number of observations to those participants who expressed at least a confidence level of six regarding the truthfulness of their answers to the well-being questions (and skips subjects' level of confidence). The reason to look only at these observations is the following: During the experiment, I realized that some non-German subjects had difficulties to fully understand the meaning of all well-being questions. This might bias the results of the well-being questions for these subjects.[6]

[6]Overall, 9 out of 102 subjects stated a confidence level below 6. One subject chose a level of 2, another subject a level of 3, three subjects chose a level of 4 and four subjects chose a level of 5. 13 subjects chose a confidence level of 6, 26 a level of 7, 40 a level of 8, and 14 a level of 9. Evidence in favor of a language problem comes from the following observation: Overall, 18 out of 102 subjects (17%) were not Germans,

Specification 3)*All - Demo.* uses the full set of available demographic controls as in Table A.3b. Specification 4)*Conf6 - Demo.* again restricts the regressions with additional controls to those subjects with a confidence level equal or above six. Specification 5) *All - Big 5* adds additionnally the Big Five Inventory (as in Table A.3c) and specification 6) *Conf6 - Big 5* restricts these regressions to those subjects with a confidence level equal or above six.

What is the overall message of Table A.4a-A.4f? First, let's focus on giving, trusting and cooperating: For SWB, KE found in their analysis a weak connection with giving. In my setting, only some long-run SWB measures become occasionally significant in some situations and some specifications. However, there is no single SWB measure which is robustly related across specifications and games, which reflects that the relationship between SWB and virtuous behavior is already quite weak without any controls, as in Table 2.2a. For PWB, the situation is different. KE found a strong connection between PWB and giving. In my results, two out of three PWB measures seem to be more or less related with virtuous behavior across games and across different specifications: the Self-Actualization Index (SAI) and Social Well-Being (SoWB). Both measures are more or less significant in three out of four expected situations (generosity in the DG, trust and cooperation in the SPD, punishment in the DG-P). Including the Big Five Inventory (specifications 5 and 6) seems only to worsen significance levels in the DG, reflecting the results of the aggregate regression analysis of Table A.3c. The Index of PWB (PWBI), however, is not robustly related with virtuous behavior which reflects the smaller significance level of this measure in Table 2.2a-2.2b.

In sum, the fairly strong connection of PWB and virtuous behavior seems to be fairly but not perfectly robust against including additional controls (when abstracting from the weakness of the PWBI and when abstracting from the weakness of the SAI (SoWB) in the trust (DG-P punishment) setting). So overall, the significance of the aggregate PWB measure in Table A.3a-A.3c seems to be driven mainly by the SAI and SoWB, not by the PWBI. The fairly weak connection of SWB and virtuous behavior disappears when controlling for additional variables. The only robust exception for SWB measures is the change in mood (MID) when cooperating or not in the SPD. The change in current happiness is not significant, but the relevant Trustor and Cooperator dummies at least have the expected sign. Overall, the change in mood may be the more reliable measure. For the change in happiness, only a single question is asked again whereas for the change in mood several questions are asked. Subjects may remember the happiness question and try to induce some artificial consistency.

Table A.4a-A.4f provide the same summary of regressions for the punishment settings:

but 5 out of 9 subjects (56%) with a confidence level below 6 were foreigners. This proportion falls to 23% (15%) foreingers for subjects with confidence level 6 (7). For reasons of consistency, all subjects with a confidence level below six are excluded and not only foreigners.

Mini-UG, SPD-P and DG-P. Already Table 2.2b showed a less clear relationship compared to Table 2.2a. When controlling for various factors, this relationship becomes even weaker. Nonetheless, we can observe the following: If there is any significant relationship for the Mini-UG (Table A.4d) at all, positive affect (PAS and partly PA) seems to be negatively related with rejecting in the DG (as expected). Those who have experienced a lower degree of good emotions in the past seem to reject more. Additionally, for Lowest Happiness (LH) the Rejector dummy is not significant but has a negative sign (as expected). For the SPD-P, we observe a shift: The signs of these two measures changes. In SPD-P, positive affect (at least the PAS) is positively related with punishing whereas LH is negatively related. Those who have experienced a higher degree of good emotions in the past seem to punish more. For the DG-P, this observed shift is even stronger to the degree that not only the PAS seems to be positively correlated with punishing but also one measure of PWB: the SAI. Hence, only for the most altruistic form of punishment, we observe that (at least) one measure of PWB is significant.

Another interesting feature of the results is that the Negative Affect Schedule (NAS) is positively correlated with punishment in all three settings. This correlation is however only significant for the punisher dummy in the SPD-P. For the SPD-P, both the Positive Affect Schedule (PAS) and the Negative Affect Schedule (NAS) are positively related with punishment free-riders. This might be an indication that different behavioral forces, one related to experiencing good feelings and one related to experiencing bad feelings, are important in this game. In sum, although results are not very robust, even these regression summaries suggest that there seems to be at least a shift from a weak negative correlation between rejecting and well-being in the Mini-UG to a weak positive correlation of punishing and well-being in the DG-P. Additionally, measures of positive and negative affect seem to play a prominent role in punishment. The aggregate analysis of Table A.3a-A.3c is not able to detect this because of aggregating the affective and the cognitive-evaluative component of SWB.

Overall, the non-aggregate analysis extends the aggregate analysis of Table A.3a-A.3c. Virtuous behavior seems to be fairly robustly correlated with the SAI and SoWB. For punishment, the affective component of SWB seems to play a crucial role and we observe an interesting shift between SP- and TP-punishment. People with high positive affect (measured by the PAS) are less likely to reject an unfair offer in the ultimatum game, but they are more likely to punish unfair behavior in the third-party punishment setting.

Table A.4a: Summary of Ordered Logit Results Coef. for the Giver Dummy in the DG

Dep. var.	Giver Dummy																					
(WB)	*1) All*			*2) Conf6*			*3) All - Demo.*			*4) Conf6 - Demo.*			*5) All - Big 5*			*6) Conf6 - Big 5*					
		p-value			p-value			p-value			p-value			p-value			p-value				
SWB																						
Long-run																						
OH	0.03	0.95		0.00	1.00		-0.01	0.98		-0.04	0.92		-0.23	0.62		-0.29	0.54				
PA	-0.03	0.93		0.10	0.81		-0.35	0.41		-0.37	0.41		-0.31	0.49		-0.39	0.40				
PAS	0.41	0.27		0.61	0.11		0.20	0.60		0.40	0.32		0.03	0.95		0.16	0.71				
NAS	0.72	**0.06**	*	0.63	0.12		0.42	0.31		0.28	0.51		0.31	0.48		0.10	0.83				
Highest/ lowest																						
HH	0.31	0.42		0.29	0.46		0.16	0.69		0.13	0.76		0.09	0.84		0.08	0.85				
LH	-0.56	0.14		-0.42	0.28		-0.18	0.66		0.03	0.95		-0.10	0.82		0.22	0.63				
Short-run																						
NH	0.09	0.81		0.15	0.71		0.12	0.77		0.26	0.54		0.09	0.84		0.28	0.52				
MI	0.02	0.95		0.11	0.78		0.04	0.91		0.21	0.62		0.10	0.81		0.26	0.54				
NHD	0.19	0.63		0.14	0.71		0.19	0.66		0.36	0.42		-0.01	0.98		0.21	0.66				
MID	0.05	0.90		0.14	0.71		0.02	0.96		0.18	0.67		-0.09	0.83		0.17	0.70				
Life satisfaction																						
SWL	0.02	0.95		0.10	0.81		0.13	0.74		0.25	0.55		-0.02	0.97		0.11	0.80				
PWB																						
PWBI	0.33	0.38		0.23	0.55		0.20	0.62		0.01	0.97		-0.28	0.52		-0.64	0.16				
SAI	0.84	**0.02**	**	0.80	**0.03**	**	0.87	**0.03**	**	0.88	**0.03**	**	0.83	**0.06**	*	0.81	**0.06**	*			
SoWB	0.76	**0.05**	**	0.70	**0.08**	*	0.73	**0.08**	*	0.73	**0.09**	*	0.63	0.14		0.52	0.24				

Notes: */** indicate significane at 10/5% level; All includes all observations, Conf6 includes only those subjects who indicated a confidence level of at least 6; Specifications 3) and 4) include additional demographic controls (age, age squared, sex, health dummy, student's subject dummy, partnership dummy, religion dummy, ethnicity); Specificantions 5) and 6) additionally include the Big Five Inventory as a control.

Table A.4b: Summary of Ordered Logit Results Coef. for the Trust Dummy in the SPD

Dep. var.	Trustor Dummy																					
(WB)	*1) All*			*2) Conf6*			*3) All - Demo.*			*4) Conf6 - Demo.*			*5) All - Big 5*			*6) Conf6 - Big 5*					
		p-value			p-value			p-value			p-value			p-value			p-value				
SWB																						
Long-run																						
OH	0.60	0.13		0.63	0.19		0.82	**0.08**	*	0.71	0.16		0.77	0.14		0.76	0.18				
PA	-0.10	0.93		0.28	0.56		-0.33	0.49		0.08	0.86		-0.36	0.49		-0.08	0.88				
PAS	0.53	0.24		0.88	0.08		0.31	0.51		0.65	0.20		0.11	0.83		0.39	0.46				
NAS	-0.43	0.18		-0.25	0.58		-0.66	0.15		-0.27	0.57		-1.00	**0.04**	**	-0.79	0.13				
Highest/ lowest																						
HH	0.31	0.43		0.36	0.46		0.30	0.54		0.30	0.55		0.44	0.41		0.37	0.50				
LH	0.38	0.63		0.42	0.37		0.73	0.12		0.46	0.35		0.85	**0.09**	*	0.76	0.15				
Short-run																						
NH	-0.19	0.87		-0.19	0.69		-0.12	0.80		-0.11	0.83		-0.11	0.81		0.02	0.97				
MI	-0.65	0.20		-0.52	0.28		-0.55	0.23		-0.51	0.30		-0.71	0.12		-0.44	0.38				
NHD	0.12	0.69		0.29	0.50		0.08	0.86		0.22	0.64		0.03	0.95		0.02	0.97				
MID	0.35	0.16		0.29	0.50		0.26	0.54		0.22	0.62		0.43	0.35		0.37	0.45				
Life satisfaction																						
SWL	-0.22	0.92		-0.17	0.70		-0.27	0.55		-0.32	0.50		-1.02	**0.04**	**	-0.96	**0.07**	*			
PWB																						
PWBI	0.72	0.21		1.08	**0.03**	**	0.68	0.14		1.04	**0.04**	**	0.32	0.53		0.71	0.18				
SAI	0.57	0.12		0.72	0.10		0.48	0.26		0.63	0.17		0.48	0.30		0.79	0.11				
SoWB	1.04	**0.00**	**	1.34	**0.00**	**	1.05	**0.02**	**	1.34	**0.01**	**	1.17	**0.02**	**	1.33	**0.01**	**			

Notes: */** indicate significane at 10/5% level; All includes all observations, Conf6 includes only those subjects who indicated a confidence level of at least 6; Specifications 3) and 4) include additional demographic controls (age, age squared, sex, health dummy, student's subject dummy, partnership dummy, religion dummy, ethnicity); Specificantions 5) and 6) additionally include the Big Five Inventory as a control.

Table A.4c: Summary of Ordered Logit Results Coef. for the Coop. Dummy in the SPD

Dep. var.	Cooperator Dummy																					
(WB)	*1) All*			*2) Conf6*			*3) All - Cont.*			*4) Conf6 - Cont.*			*5) All - Big 5*			*6) Conf6 - Big 5*					
		p-value			p-value			p-value			p-value			p-value			p-value				
SWB																						
Long-run																						
OH	-0.08	0.86		-0.19	0.68		-0.25	0.61		-0.38	0.46		-0.33	0.53		-0.59	0.29				
PA	0.68	0.12		0.89	**0.05**	*	0.50	0.29		0.83	0.10		0.64	0.22		0.75	0.18				
PAS	0.53	0.21		0.74	**0.10**	*	0.32	0.47		0.59	0.20		0.19	0.70		0.34	0.50				
NAS	-0.12	0.76		-0.16	0.71		-0.11	0.81		-0.20	0.66		-0.44	0.35		-0.71	0.16				
Highest/ lowest																						
HH	0.23	0.59		0.39	0.38		0.06	0.90		0.23	0.64		-0.19	0.70		-0.06	0.91				
LH	-0.60	0.16		-0.40	0.36		-0.71	0.14		-0.40	0.42		-0.69	0.17		-0.32	0.54				
Short-run																						
NH	0.34	0.45		0.45	0.33		0.49	0.32		0.70	0.17		0.53	0.30		0.78	0.15				
MI	-0.35	0.42		-0.04	0.94		-0.41	0.38		0.02	0.96		-0.46	0.34		0.04	0.93				
NHD	0.59	0.19		0.45	0.33		0.64	0.18		0.64	0.21		0.53	0.29		0.48	0.37				
MID	1.21	**0.01**	**	1.08	**0.02**	**	1.32	**0.01**	**	1.28	**0.01**	**	1.48	**0.00**	**	1.56	**0.00**	**			
Life satisfaction																						
SWL	-0.06	0.89		-0.01	0.98		-0.08	0.87		0.07	0.89		-0.19	0.70		-0.16	0.77				
PWB																						
PWBI	0.26	0.54		0.30	0.49		0.21	0.64		0.22	0.63		-0.18	0.70		-0.45	0.37				
SAI	1.12	**0.01**	**	1.15	**0.01**	**	1.08	**0.02**	**	1.21	**0.01**	**	1.16	**0.02**	**	1.37	**0.01**	**			
SoWB	1.61	**0.00**	**	1.50	**0.00**	**	1.53	**0.00**	**	1.43	**0.00**	**	1.67	**0.00**	**	1.32	**0.02**	**			

Notes: */** indicate significane at 10/5% level; All includes all observations, Conf6 includes only those subjects who indicated a confidence level of at least 6; Specifications 3) and 4) include additional demographic controls (age, age squared, sex, health dummy, student's subject dummy, partnership dummy, religion dummy, ethnicity); Specificantions 5) and 6) additionally include the Big Five Inventory as a control.

Table A.4d: Summary of Ordered Logit Results coef. for the Rejection Dummy in the Mini-UG

Dep. var.	Rejector Dummy															
(WB)	*1) All*			*2) Conf6*			*3) All - Cont.*			*4) Conf6 - Cont.*			*5) All - Big 5*			*6) Conf6 - Big 5*		
		p-value			p-value			p-value			p-value			p-value			p-value	
SWB																		
Long-run																		
OH	0.03	0.95		-0.02	0.95		0.05	0.89		0.00	0.99		0.17	0.68		0.14	0.76	
PA	-0.72	**0.06**	*	-0.54	0.17		-0.71	**0.07**	*	-0.54	0.19		-0.57	0.15		-0.39	0.37	
PAS	-0.73	**0.04**	**	-0.63	**0.09**	*	-0.76	**0.04**	**	-0.63	0.10		-0.77	**0.04**	**	-0.55	0.16	
NAS	0.30	0.39		0.36	0.33		0.29	0.44		0.33	0.40		0.10	0.80		0.14	0.74	
Highest/ lowest																		
HH	-0.13	0.73		-0.29	0.47		-0.20	0.61		-0.40	0.33		0.02	0.97		-0.19	0.66	
LH	0.25	0.50		0.14	0.71		0.29	0.43		0.19	0.62		0.56	0.15		0.40	0.32	
Short-run																		
NH	-0.24	0.52		-0.24	0.53		-0.33	0.37		-0.32	0.42		-0.37	0.33		-0.30	0.47	
MI	-0.31	0.37		-0.20	0.60		-0.32	0.39		-0.16	0.68		-0.13	0.72		0.10	0.80	
Life satisfaction																		
SWL	-0.28	0.44		-0.41	0.29		-0.33	0.38		-0.45	0.26		-0.17	0.65		-0.21	0.60	
PWB																		
PWBI	-0.47	0.19		-0.54	0.16		-0.51	0.16		-0.60	0.13		-0.24	0.52		-0.13	0.75	
SAI	0.06	0.86		-0.16	0.66		0.00	0.99		-0.20	0.60		0.14	0.70		-0.04	0.92	
SoWB	-0.19	0.59		0.03	0.94		-0.21	0.57		0.06	0.88		-0.08	0.84		0.25	0.52	

Notes: */** indicate significane at 10/5% level; All includes all observations, Conf6 includes only those subjects who indicated a confidence level of at least 6; Specifications 3) and 4) include additional demographic controls (age, age squared, sex, health dummy, student's subject dummy, partnership dummy, religion dummy, ethnicity); Specificantions 5) and 6) additionally include the Big Five Inventory as a control.

Table A.4e: Summary of Ordered Logit Results coef. for the Punishment Dummy in the SPD-P

Dep. var.	Punisher Dummy																	
(WB)	*1) All*			*2) Conf6*			*3) All - Cont.*			*4) Conf6 - Cont.*			*5) All - Big 5*			*6) Conf6 - Big 5*		
		p-value			p-value			p-value			p-value			p-value			p-value	
SWB																		
Long-run																		
OH	0.00	0.99		0.14	0.73		-0.08	0.83		0.13	0.76		-0.24	0.56		-0.10	0.82	
PA	0.00	0.99		-0.17	0.66		-0.06	0.87		-0.31	0.44		-0.26	0.51		-0.65	0.13	
PAS	0.66	**0.08**	*	0.75	**0.05**	**	0.64	**0.08**	*	0.77	**0.05**	**	0.65	**0.09**	*	0.64	0.12	
NAS	0.60	**0.10**	*	0.62	**0.10**	*	0.64	**0.09**	*	0.60	0.13		0.67	**0.08**	*	0.77	**0.06**	*
Highest/ lowest																		
HH	0.00	1.00		0.07	0.87		0.03	0.94		0.05	0.91		0.04	0.93		-0.02	0.96	
LH	-0.62	**0.10**	*	-0.61	0.12		-0.84	**0.03**	**	-0.74	**0.08**	*	-0.93	**0.02**	**	-0.87	**0.05**	**
Short-run																		
NH	-0.37	0.31		-0.44	0.25		-0.47	0.22		-0.59	0.15		-0.44	0.25		-0.66	0.11	
MI	-0.17	0.63		-0.21	0.57		-0.26	0.48		-0.33	0.40		-0.37	0.32		-0.52	0.19	
Life satisfaction																		
SWL	0.11	0.76		0.22	0.55		0.07	0.85		0.23	0.55		-0.23	0.55		-0.16	0.69	
PWB																		
PWBI	-0.05	0.89		0.06	0.86		-0.18	0.63		-0.04	0.91		-0.51	0.19		-0.79	**0.06**	*
SAI	0.40	0.27		0.34	0.36		0.39	0.30		0.32	0.41		0.39	0.31		0.27	0.50	
SoWB	0.41	0.26		0.31	0.41		0.29	0.44		0.18	0.65		0.11	0.78		-0.23	0.59	

Notes: */** indicate significane at 10/5% level; All includes all observations, Conf6 includes only those subjects who indicated a confidence level of at least 6; Specifications 3) and 4) include additional demographic controls (age, age squared, sex, health dummy, student's subject dummy, partnership dummy, religion dummy, ethnicity); Specificantions 5) and 6) additionally include the Big Five Inventory as a control.

Table A.4f: Summary of Ordered Logit Results Coef. for the Punishment Dummy in the DG-P

Dep. var.	Punisher Dummy																	
(WB)	*1) All*			*2) Conf6*			*3) All - Cont.*			*4) Conf6 - Cont.*			*5) All - Big 5*			*6) Conf6 - Big 5*		
		p-value			p-value			p-value			p-value			p-value			p-value	
SWB																		
Long-run																		
OH	-0.08	0.86		0.00	1.00		-0.25	0.58		-0.15	0.75		-0.39	0.42		-0.32	0.53	
PA	-0.05	0.91		-0.04	0.92		-0.08	0.86		-0.18	0.69		0.22	0.63		0.06	0.90	
PAS	0.70	**0.09**	*	0.76	**0.07**	*	0.68	0.11		0.83	**0.06**	*	0.76	**0.07**	*	0.79	**0.08**	*
NAS	0.28	0.49		0.41	0.32		0.54	0.20		0.62	0.15		0.49	0.25		0.71	0.11	
Highest/ lowest																		
HH	0.05	0.91		0.02	0.97		0.11	0.82		0.01	0.98		0.48	0.33		0.33	0.51	
LH	-0.16	0.70		-0.25	0.56		-0.28	0.51		-0.36	0.43		-0.18	0.69		-0.22	0.64	
Short-run																		
NH	-0.20	0.66		-0.33	0.46		-0.20	0.67		-0.35	0.45		-0.11	0.82		-0.30	0.52	
MI	-0.02	0.96		-0.08	0.84		-0.04	0.92		-0.11	0.81		0.12	0.78		0.06	0.89	
Life satisfaction																		
SWL	0.32	0.45		0.47	0.28		0.21	0.63		0.35	0.45		-0.08	0.85		0.02	0.97	
PWB																		
PWBI	0.44	0.30		0.61	0.16		0.24	0.58		0.37	0.41		-0.06	0.90		-0.20	0.67	
SAI	1.10	**0.01**	**	1.18	**0.01**	**	1.07	**0.01**	**	1.15	**0.01**	**	1.09	**0.01**	**	1.18	**0.01**	**
SoWB	0.55	0.20		0.72	0.10		0.44	0.31		0.61	0.18		0.35	0.45		0.32	0.50	

Notes: */** indicate significane at 10/5% level; All includes all observations, Conf6 includes only those subjects who indicated a confidence level of at least 6; Specifications 3) and 4) include additional demographic controls (age, age squared, sex, health dummy, student's subject dummy, partnership dummy, religion dummy, ethnicity); Specificantions 5) and 6) additionally include the Big Five Inventory as a control.

A.2.4 Additional Analysis of the Virtuousness Hypothesis

In this section, a more detailed analysis of the virtuousness hypothesis is presented. A crucial aspect, not discussed in the main text, is that Now Happiness (NH) and the Mood Index (MI) were asked directly before and after the SPD. This means that subjects made both decisions as first and second mover in the SPD before answering these well-being questions again. I hence cannot clearly distinguish whether trusting or cooperating leads to a difference in the NHD and MID. However, the Cooperator dummy performs better than the Trustor dummy in the regression analysis of Appendix A.2.2 (Table A.3a-A.3c). The reason why this is the case can be seen from Table A.5. This table presents the Now Happiness Difference (NID) and the Mood Index Difference (MID) for 1) Givers and Nongivers, 2) for Trutors and Nontrustors, 3) for Cooperators and Free-Riders, and 4) for those who either trusted or cooperated versus those who did not behave virtuously at all. Additionally, the table presents the data separately for the two different sequences (DG-SPD vs. SPD-DG) I employed.

Looking first at the pooled data for the SPD, one can see that the NHD and MID are very similar for Trustors, Cooperators and subjects who trust or cooperate. On the contrary, Nontrustors experience a decrease in happiness and mood but this decrease is not as big as for Free-riders or subjects who are both Free-riders and Nontrustors (third column). A plausible explanation for this result is the following: Only 64% of subjects trusted, but 77% of subjects cooperated. Most subjects who trusted also cooperated (61 out of 65), but more people who cooperated did not trust (18 out of 79). Hence, many of the Nontrustors cooperated (18 out 37). However, out of the Free-riders a smaller proportion trusted (4 out of 23). This means most Free-riders did not behave virtuously at all. This seems to be the reason why Free-riders and those who are both Free-riders and Nontrustors suffer from a similiar decrease in current happiness and mood.

There are two plausible mechanisms that I cannot distinguish: Either not behaving virtuously at all makes people unhappy or free-riding has much worse effects than not trusting. The problem is that I basically only observe free-riding subjects who also do not trust. A possible reasoning for the first line of argument is provided by Becchetti and Antoni (2010). They observe that in a standard trust game only trustors experience gains in short-run happiness but not trustees because only trustors have a total payoff enhancing power. In my SPD, sent money is doubled for both the first and the second mover. Not using this efficiency enhancing power in both possible situations might decrease short-run happiness. Subjects may feel guilty because of not fulfilling the expectations of the other player (see Charness and Dufwenberg, 2006). However, it might still be true that free-riding itself makes people unhappy because returning a favor (cooperating) might be considered to be a social norm whereas trusting might not be considered such a norm. This would potentially make not cooperating more harmful for well-being than not trusting.

Table A.5: Now Happiness and Mood Index Difference- DG and SPD

1) Dictator Game

Sequence		Nongivers	Givers	T-Test
Pooled	NHD	0.00	0.10	0.35
	MID	-1.26	-0.94	0.42
DG-SPD	NHD	0.20	0.18	0.52
	MID	-1.00	-1.82	0.62
SPD-DG	NHD	-0.21	0.00	0.23
	MID	-1.53	0.10	**0.10**

2) Sequential Prisoner's Dilemma

Sequence		Nontrustors	Trustors	T-Test
Pooled	NHD	-0.16	0.06	0.16
	MID	-3.00	0.37	**0.02**
SPD-DG	NHD	0.00	-0.03	0.54
	MID	-2.06	-1.97	0.49
DG-SPD	NHD	-0.32	0.14	**0.07**
	MID	-3.89	2.37	**0.00**

3) Sequential Prisoner's Dilemma

Sequence		Free-riders	Cooperators	T-Test
Pooled	NHD	-0.43	0.10	**0.02**
	MID	-5.48	0.49	**0.00**
SPD-DG	NHD	-0.22	0.03	0.28
	MID	-6.67	-0.92	**0.03**
DG-SPD	NHD	-0.57	0.18	**0.01**
	MID	-4.71	1.88	**0.00**

4) Sequential Prisoner's Dilemma

Sequence		Nontrustors & Free-riders	Trustor or Cooperators	T-Test
Pooled	NHD	-0.42	0.07	**0.04**
	MID	-5.95	0.31	**0.00**
SPD-DG	NHD	0.00	-0.02	0.52
	MID	-6.29	-1.27	**0.07**
DG-SPD	NHD	-0.67	0.17	**0.01**
	MID	-5.75	1.86	**0.00**

Notes: The T-Test column provides p-values according to one-tail t-tests of the null hypothesis that subjects who behave virtuously experience a higher NHD and MID. All p-values below 10% are in bold type, n=102.

Table A.5 additionally presents the NHD and the MID separately for the two different sequences (DG-SPD vs. SPD-DG) I employed. For the SPD, the difference in the NID and the MID between virtuously and non-virtuously behaving subjects is more pronounced when the SPD is played as a second game. In the DG-SPD sequence, Trustors and Cooperators seem to experience a higher NID/MID than in the SPD-DG sequence. One plausible explanation for this observation is that playing the DG first highlights the efficiency enhancing feature of the SPD, leading to an increase in happiness and mood as observed by Becchetti and Antoni (2010). For the DG, I observe a marginally significant

difference between Nongivers and Givers when the SPD is played first. One - fairly speculative - explanation might be that playing a game with potentially high payoffs first increases average payoff expectations. Givers may be aware of this and experience a mood increase when meeting the expectations of others (and avoiding guilt) by giving in the DG. The same might be true for Nongivers in the other direction.

A.3 Items on Well-Being Measures

This section presents the items used in my well-being questionnaire. Nearly all items are also used by KE, only the Social Well-Being (SoWB) Scale is added.

Subjective Well-Being

Long-run Happiness

- **Overall Happiness (OH)**
  A 9-point-scale ranging from "extremely unhappy" to "extremly happy" is used:
  OH: *Overall*, how would you describe yourself?

- **Bradburn's (1969) Positive Affect (PA) and Negative Affect (NA) Scales**
  Subjects have to answer with "YES" or "NO" for each item:
  During the past few weeks did you ever feel...
  PA items:
  - Pleased about having accomplished something?
  - That things were going your way?
  - Proud because someone complimented you on something you had done?
  - Particularly excited or interested in something?
  - On top of the world?

  NA items:
  - Depressed or very unhappy?
  - Very lonely or remote from other people?
  - Upset because someone criticized you?
  - So restless that you couldn't sit long in a chair?
  - Bored?

- **Watson, Clark and Tellegen's (1988) Positive Affect (PAS) and Negative Affect (NAS) Schedules**
  Subjects use a 5-point-scale ranging from "very slightly or not at all" to "extremely" to indicate to what extent they feel the way the words suggest:
  - **PAS items**: interested, alert, excited, inspired, strong, determined, attentive, active, enthusiastic, proud
  - **NAS items**: irritable, distressed, ashamed, upset, nervous, guilty, scared, jittery, hostile, afraid

Highest/ Lowest Happiness

- **Highest/ Lowest Happiness (HH/ LH)**
  A 9-point-scale ranging from "extremely unhappy" to "extremly happy" is used:

  HH: Over the past week, what is the *highest* level you experienced?
  LH: Over the past week, what is the *lowest* level you experienced?

Short-run Happiness

- **Now Happiness (NH)**
  A 9-point-scale ranging from "extremely unhappy" to "extremly happy" is used:
  NH: *Right now*, how would you describe yourself?

- **Batson, et al.'s (1988) Mood Index (MI)**
  On a 9-point-scale, subjects have to express their current mood for several pairs of adjectives

  - **Mood items**: bad mood-good mood, sad-happy, depressed-elated, dissatisfied-satisfied, gloomy-cheerful, displeased-pleased, sorrowful-joyful
  - **Fillers**: nervous-calm, tense-relaxed, uncomfortable-comfortable, apathetic-caring, lethargic-energetic, unconfident-confident, unresponsive-emotional, passive-active

Life Satisfaction

- **Diener, et al.'s (1985) Satisfaction with Life Scale (SWL)**
  A 7-point-scale ranging from "strongly disagree" to "strongly agree" is used:

  1. In most ways my life is close to my ideal.
  2. The conditions of my life are excellent.
  3. I am satisfied with my life.
  4. So far I have gotten the important things I want in life.
  5. If I could live my life over, I would change almost nothing.

Psychological Well-Being

- **Ryff's (1995) Scales of Psychological Well-Being (SPWB)**
  A 6-point-scale ranging from "strongly disagree" to "strongly agree" is used:

  1. I tend to be influenced by people with strong opinions.*
  2. In general, I feel I am in charge of the situation in which I live.
  3. I think it is important to have new experiences that challenge how you think about yourself and the world.
  4. Maintaining close relationships has been difficult and frustrating for me.*
  5. I live life one day at a time and don't really think about the future.*
  6. When I look at the story of my life, I am pleased with how things have turned out.
  7. I have confidence in my opinions, even if they are contrary to the general consensus.|
  8. The demands of everyday life often get me down.*
  9. For me, life has been a continuous process of learning, changing and growth.|
  10. People would describe me as a giving person, willing to share my time with others.
  11. Some people wander aimlessly through life, but I am not one of them.|
  12. I like most aspects of my personality.
  13. I judge myself by what I think is important, not by the values of what others think is important.
  14. I am quite good at managing the many responsibilities of my daily life.|
  15. I gave up trying to make a big improvements or changes in my life a long time ago.*
  16. I have not experienced many warm and trusting relationships with others.*|
  17. I sometimes feel as if I've done all there is to do in life.*
  18. In many ways, I feel disappointed about my achievements in life.*|

  * indicates reverse scored items. | indicates items selected for Index of PWB (PWBI)

- **Jones and Crandall's (1986) Self-Actualization Index (SAI)**
  A 4-point-scale ranging from "disagree" to "agree" is used:

  1. I do not feel ashamed of any of my emotions.
  2. I feel I must do what others expect me to do.*
  3. I believe that people are essentially good and can be trusted.
  4. I feel free to be angry at those I love.
  5. It is always necessary that others approve of what I do.*

6. I don't accept my own weaknesses.*
7. I can like people without having to approve of them.
8. I fear failure.*
9. I avoid attempts to analyze and simplify complex domains.*
10. It is better to be yourself than to be popular.
11. I have no mission in life to which I feel especially dedicated.*
12. I can express my feelings even when they may result in undesirable consequences.
13. I do not feel responsible to help anybody.*
14. I am bothered by fears of being inadequate.*
15. I am loved because I give love.

* indicates reverse-scored items.

- **Keyes' (1998) Social Well-Being (SoWB)**
  A 6-point-scale ranging from "strongly disagree" to "strongly agree" is used:

1. The world is too complex for me.*
2. I don't feel I belong to anything I'd call a community.*
3. People who do a favor expect nothing in return.
4. I have something valuable to give to the world.
5. The world is becoming a better place for everyone.
6. I feel close to other people in my community.
7. My daily activities do not produce anything worthwhile for my community.*
8. I cannot make sense of what's going on in the world.*
9. Society has stopped making progress.
10. People do not care about other people's problems.*
11. My community is a source of comfort.
12. I find it easy to predict what will happen next in society.
13. Society isn't improving for people like me.*
14. I believe that people are kind.
15. I nothing important to contribute to society.*

* indicates reverse-scored items.

A.4 Instructions

This section provides the instructions for all six games (DG, SPD, JOD-G, Mini-UG, SPD-P, DG-P), translated from the German original. Additionally, for the more complicated games (SPD, SPD-P, DG-P), control questions were asked that are provided at the end of this section. The general structure/ design of the instructions is based on those of Blanco et al. (2011). For each individual game, instructions are also partially based on those papers from which the game was taken from (cf. Section 2.3.2).

Instructions - DG (Dictator Game)

In this section of the experiment the **situation** is as follows:
At the beginning, Person A gets 20 EUR. Person B gets 0 EUR. Then, Person A has the opportunity to send money to Person B. However, Person A is not obliged to send money to Person B.

The Details:

So **Person A** gets **20 EUR** and has the following two alternatives: Sending no money to B (**0 EUR**) or sending money to B (either **2, 4, 6, 8, 10, 12, 14, 16, 18, or 20 EUR**).

Person B gets **0 EUR** at the beginning and knows that Person A has the described alternatives. Although B will be informed about A's decision, B does not make any decision him- or herself and has to accept A's decision.

Hence, we have the following **payoffs**:
For **Person A**: 20 EUR - money send to B
For **Person B**: Money send from B

Four examples:

1. If A sends **0 EUR**, A's payoff is 20 EUR and B' payoff is 0 EUR.
2. If A sends **10 EUR**, A's payoff is 10 EUR and B' payoff is 10 EUR.
3. If A sends **20 EUR**, A's payoff is 0 EUR and B' payoff is 20 EUR.
4. If A sends **4 EUR**, A's payoff is 16 EUR and B' payoff is 4 EUR.

Your decisions:

In the following, you will have to decide in the role of Person A.
If the computer chooses this section for payment, the computer will randomly match you with another participant. Additionally, it will be randomly determined who will be assigned to which role. For the payment of **both** participants, only the decision of the participant who is assigned to the role of Person A will be of interest.

Instructions - SPD (Sequential Prisoner's Dilemma)

In this section of the experiment the **situation** is as follows:
Person A and Person B both have an endowment of 10 EUR at the beginning and they have to decide how to use their endowment. Both can either keep their money or send it to the other person. If money is sent, this money is doubled by the computer.

The Details:

This section of the experiment consists of two consecutive stages:
In the **first stage**, Person A has two alternatives: sending **0 EUR** or **10 EUR** to Person B. In doing so, sent money is doubled. This decision determines how much money Person B has at the beginning of the second stage. We have two cases:

1. If A sends **10 EUR**, this amount is doubled. Because B already owns 10 EUR, B has 30 EUR in total now (and A has 0 EUR).
2. If A sends **0 EUR**, B does not get any additional money and owns his or her initial endowment of 10 EUR (as A does).

In the **second stage**, Person B has the following six different alternatives: sending **0, 2, 4, 6, 8 or 10 EUR** back to Person A. The amount that B chooses is again doubled. Person B can make his or her decision conditional on A's choice.

In **case 1)**, Person B owns 30 EUR at the beginning of the second stage whereas B owns 10 EUR in **case 2)**. For calculating B's final payoff, one still has to subtract the amount sent by B. Because the amount sent by B is also doubled, A can earn between **0 and 20 EUR** in this stage. In case 2), Person A gets an additional 10 EUR form stage 1, in case 1) A gets no additional money.

Two examples:

1. A sends **10 EUR**. Hence for a start, B has 30 EUR. If B also sends **10 EUR**, B finally earns 20 EUR, as does A. If B on the contrary sends 0 EUR, B finally earns 30 EUR and A earns 0 EUR.

2. A sends **0 EUR**. Hence for a start, B has 10 EUR. If B also sends **0 EUR**, B finally earns 10 EUR, as does A. If B on the contrary sends 10 EUR, B finally earns 0 EUR and A earns 30 EUR.

Mathematically we can express these **payoffs** as follows:

For **Person A**: 10 EUR - transfer to B + 2 * transfer from B

For **Person B**: 10 EUR - transfer to A + 2 * transfer from A

For further calculations, you can use the implemented payoff calculator

Your decisions:

In the following, you will have to decide both in the role of Person A as well as in the role of Person B. In the role of Person B, you will have to make a decision for both possible alternatives of Person A (0 EUR vs. 10 EUR).

If the computer chooses this section for payment, the computer will randomly match you with another participant. Additionally, it will be randomly determined who will be assigned to which role. For your payment, only your decision in this role (either as Person A or B) will be important.

Instructions - JOD-G (Joy-of-Destruction Game)

In this section of the experiment the **situation** is as follows:

Person A and Person B both get 11 EUR and have the opportunity to reduce the other's payoff but they do not have to do this.

The Details:

Both **Person A** and **Person B** are initially endowed with 11 EUR and then **simultaneously** make the following decision: Leaving the other person's income **unchanged** or reducing it by **1, 2, 3, 4, or 5 EUR**. Neither leaving the other's income unchanged nor reducing it costs money. **Importantly** however, only in **two out of three** cases A's and B's decision will really determine payoffs. In **one out of three** cases these decisions are irrelevant and the computer will "overwrite" them by reducing both incomes either by **0, 1, 2, 3, 4, or 5 EUR**, where all amounts are equally likely.

Importantly, neither Person A nor Person B will be informed about which case has occurred. If for example own income is reduced, it is not evident whether the other person **or** the computer is responsible for this reduction.

Three examples:

1. If both A and B decide to leave the other's payoff unchanged and if the computer does not alter this, both A and B finally earn **11 EUR**. If the computer overwrites these decisions and chooses 2 EUR as a reduction, both A and B finally earn **9 EUR**.

2. If both A and B decide to reduce the other's payoff by 5 EUR and if the computer does not alter this, both A and B finally earn **6 EUR**.

3. If A decides to reduce B's payoff by 1 EUR and if B leaves A's payoff unchanged, A finally earns **11 EUR** and B earns **10 EUR** (in case the computer does not alter these decisions).

Your decisions:

Because Person A and B have symmetric roles, you will only have to decide on the next screen whether you want to reduce your counterpart's payoff or whether you want to leave it unchanged. If the computer chooses this section for payment, the computer will randomly match you with another participant.

Instructions - Mini-UG (Mini Ultimatum-Game)

In this section of the experiment the **situation** is as follows:
Person A has two alternatives to split 20 EUR between him- or herself and Person B. Person B can accept or reject this choice.

The Details:

This section of the experiment consists of two consecutive stages:
In the **first stage**, Person A has to choose between two proposals on how to split the 20 EUR:

- **Proposal 1:** Person A and Person B both get **10 EUR**.
- **Proposal 2:** Person A gets **18 EUR** and Person B gets **2 EUR**.

In the **second stage**, Person B has two alternatives: B can either **accept** or **reject** A's choice. If Person B accepts A's choice, payoffs are according to the chosen proposal. If Person B rejects A's choice, both A and B receive **0 EUR**.

Your decisions:

You will have to decide both in the role of Person A and in the role of Person B. In the role of Person B, you will have to decide for both proposals whether you would like to accept or reject them.
If the computer chooses this section for payment, the computer will randomly match you with another participant. Additionally, it will be randomly determined who will be assigned to which role. For your payment, only the decision in the assigned role (either as Person A or B) will then be of interest.

Instructions - SPD-P (Sequential Prisoner's Dilemma with Punishment)

In this section of the experiment the **situation** is as follows:
Similar to Section 2) [SPD] Person A and Person B get 10 EUR and have to decide how to use this money. Additionally however, there is a Person C. C gets 20 EUR and has the opportunity to reduce B's payoff.

The Details:

This section of the experiment consists of three consecutive stages:
In the **first stage**, Person A has to choose between two alternatives: Sending either **0 EUR** or **10 EUR** to Person B. We have two cases:

- **Case a**: If A sends **0 EUR, all other following stages are omitted** and the final payoffs are 10 EUR both for A and B and 20 EUR for C.
- **Case b**: If A sends **10 EUR**, this money is doubled. Because B already owns 10 EUR initially, B then temporarily owns 30 (and A 0 EUR).

In the **second stage**, Person B has two alternatives: Sending either **0 EUR** or **10 EUR** to Person A. The amount that Person B chooses is again doubled by the computer.
Hence, we have the following temporary payoffs:

- **Case 1**: If B sends **0 EUR**, B gets 30 EUR and A gets 0 EUR (unchanged payoffs of case b)
- **Case 2**: If B in contrast sends **10 EUR**, both A and B get 20 EUR each.

In the **third stage**, however, Person C has the opportunity to use part of his or her endowment to reduce Person B's payoff: This is possible in steps of 50 cents: **50, 100, 150, . . . , 600 cents**. Person C can condition his or her decision on B's decision.

Fifty cents invested by C reduce B's payoff **250 cents**, or 2.50 EUR. If C e.g. invests 200 cents (in order to reduce B's payoff), Person B will loose 10 EUR compared to the second stage. B's minimal payoff, however, can in the worst case only be reduced to 0 EUR (and does not get negative). If Person C does not invest anything in reducing B's payoff, all payoffs are the same as at the end of the second stage.

Three Examples:

1. If Person A sends **0 EUR**, stages two and three are dropped. A and B get 10 EUR and C gets 20 EUR (compare **case a**)
2. If Person A sends **10 EUR**, if B sends **0 EUR**, and if C does **not** reduce, A receives 0 EUR, B 30 EUR and C 20 EUR (compare **case 1**). If in contrast C invests e.g. **400 cents** (=4 EUR), A receives 0 EUR, B 10 EUR and C 16 EUR.
3. If Person A sends **10 EUR**, if B sends **10 EUR**, and if C does **not** reduce, A receives 20 EUR, B 20 EUR and C 20 EUR (compare **case 2**). If in contrast C invests e.g. **50 cents**, A receives 20 EUR, B 17.50 EUR and C 19.50 EUR.

For further calculations, you can use the implemented payoff calculator

Your decisions:

In the following, you will have to decide in all three roles (Person A, Person B, and Person C). In the role of Person C, you will have to make decision for both alternative actions of Person B. If the computer chooses this section for payment, the computer will randomly match you with another participant. Additionally, it will be randomly determined who will be assigned to which role.

Instructions - DG-P (Dictator Game with Punishment)

In this section of the experiment the **situation** is as follows:
At the beginning, Person A and Person B get 5 EUR each. Then, Person A additionally gets 10 EUR and he or she can send money to Person B, but does not have to do this. In addition, Person C can reduce A's payoff.

The Details:

This section of the experiment consists of two consecutive stages:
In the **first stage**, **Person A** has three alternatives: Sending **0, 2.50 or 5 EUR** to Person B. **Person B** knows about Person A's alternatives but does not make any decision him- or herself and has to accept A's decision.

In the **second stage**, **Person C** can reduce Person A's payoff. C can invest either **0, 1, 2, 3, 4, or 5 EUR**. One euro invested by C reduces A's payoff exactly by one euro. Person C can condition his decision with respect to A's decision.

Hence, we have the following **payoffs**:

- **Person A**: 15 EUR - amount sent to B - reduction by C
- **Person B**: 5 EUR + amount sent by A
- **Person C**: 20 EUR - reduction of A's payoff

Two Examples:

1. If Person A sends **0 EUR** and if C does **not** reduce, we have the following payoffs: A 15 EUR, B 5 EUR, and C 20 EUR. If in contrast C invests e.g. **3 EUR**, payoffs change to: A 12 EUR, B 5 EUR, and C 17 EUR.
2. If Person A sends **5 EUR** and if C does **not** reduce, we have the following payoffs: A 10 EUR, B 10 EUR, and C 20 EUR. If in contrast C invests e.g. **1 EUR**, payoffs change to: A 9 EUR, B 10 EUR, and C 19 EUR.

Your decisions:

In the following, you will have to make deicsions in the role of Person A and C. In the role of C, you will have to decide for all three alternatives of A (0, 2.50, or 5 EUR).
If the computer chooses this section for payment, the computer will randomly match you with another participant. Additionally, it will be randomly determined who will be assigned to which role.

Control Questions (SPD, SPD-P, DG-P)

SPD - Sequential Prisoner's Dilemma

1. Assume the following: A sends 10 EUR. Hence, A **temporarily** owns 0 EUR and B owns 30 EUR. In this situation, B sends 4 EUR.
   - In this situation, how large is the payoff of Person A?
   - In this situation, how large is the payoff of Person B?
2. Assume the following: A sends 10 EUR. Hence, A **temporarily** owns 0 EUR and B owns 30 EUR. In this situation, B sends 6 EUR.
   - In this situation, how large is the payoff of Person A?
   - In this situation, how large is the payoff of Person B?

SPD-P - Sequential Prisoner's Dilemma with Punishment

1. Assume the following: Case 2 has occurred and C invests **400 cents** (also compare example 3).
   - How large is A's payoff?
   - How large is B's payoff?
   - How large is C's payoff?
2. Assume the following: Case 1 has occurred and C invests **600 cents** (also compare example 2).
   - How large is A's payoff?
   - How large is B's payoff?
   - How large is C's payoff?

DG-P - Dictator Game with Punishment

1. Assume the following: Person A sends **2.50 EUR** and C invests **5 EUR**.
   - How large is A's payoff (in EUR)?
   - How large is B's payoff (in EUR)?
   - How large is C's payoff (in EUR)?

Appendix B

Appendix to Chapter 3

In this appendix, I first provide three additional tables and one additional figure (Appendix B.1). Afterwards, I formally derive the theoretical predictions of the different hypotheses outlined in Section 3.3 using preferences of Fehr and Schmidt (1999) and the framework of Hart and Moore (2008) (Appendix B.2). Afterwards, an additional analysis regarding subjects payoffs is outlined (Appendix B.3). Finally, I provide a translated version of the experimental instructions (Appendix B.4).

B.1 Tables and Figures

Table B.1: Panel Regressions on Effort - CasRP (recession data only)

	(1) RE	(2) RE	(3) RE	(4) FE	(5) FE	(6) FE
Wage	0.105***	0.0971		0.104***	0.0895	
	(0.0114)	(0.0637)		(0.0109)	(0.0648)	
Wage cut	-0.247**	-0.241*	-0.261**	-0.305*	-0.296	-0.242**
	(0.123)	(0.140)	(0.123)	(0.135)	(0.151)	(0.0875)
Wage^2		✓			✓	
Wage-dummies			✓			✓
Period & Period^2	✓	✓	✓	✓	✓	✓
Constant	-2.024***	-1.813	7.953***	-1.978*	-1.610	8.065***
	(0.771)	(1.858)	(1.468)	(0.795)	(1.891)	(1.305)
Observations	207	207	207	207	207	207
R^2	0.579	0.579	0.602	0.578	0.577	0.589

Notes: Panel random effects (RE) and fixed effects (FE) regressions on effort for the CasRP-F treatment (only for recession data). Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table B.2: Panel Regressions on Wages

	(1) RE	(2) RE	(3) RE	(4) FE
CasRP-dummy	0.257	-1.659	-3.320**	
	(4.046)	(4.138)	(1.516)	
CasRP-F-dummy	1.388	-0.834	-3.231**	
	(3.408)	(3.825)	(1.583)	
WasRP-dummy	3.000	1.159	-1.469	
	(4.420)	(4.453)	(1.361)	
Recession-dummy	-6.832***	-11.51***	-10.61***	-11.13***
	(1.135)	(1.550)	(2.226)	(2.043)
Rec x CasRP		5.822***	6.129**	5.463**
		(2.250)	(2.415)	(2.336)
Rec x CasRP-F		7.051**	7.468**	6.949**
		(3.247)	(2.951)	(2.893)
Rec x WasRP		5.307**	4.805**	4.304*
		(2.153)	(2.261)	(2.176)
Wage_{t-1} (No-Rec)			0.373***	0.136***
			(0.0432)	(0.0258)
Wage_{t-1} (Rec)			0.405***	0.148***
			(0.0463)	(0.0262)
Effort_{t-1} (No-Rec)			3.303***	2.615***
			(0.240)	(0.214)
Effort_{t-1} (Rec)			3.393***	2.902***
			(0.293)	(0.268)
First wage			0.0730***	
			(0.0271)	
Period & Period^2	✓	✓	✓	✓
Constant	57.82***	59.45***	22.59***	40.91***
	(3.195)	(3.255)	(2.816)	(1.668)
Observations	2858	2858	2696	2696
R^2	0.0329	0.0384	0.610	0.310

Notes: Panel random effects (RE) and fixed effects (FE) regressions on wages for all treatments. Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table B.3: Panel Regressions on Effort

	(1) RE	(2) RE	(3) RE	(4) RE	(5) FE	(6) FE
Wage	0.132***	0.132***	0.132***	0.162***	0.131***	0.161***
	(0.0183)	(0.0184)	(0.0185)	(0.0341)	(0.0186)	(0.0336)
CasRP-dummy	0.219	0.345	0.343	1.217		
	(0.281)	(0.264)	(0.265)	(1.143)		
CasRP-F-dummy	0.0701	0.257	0.255	0.530		
	(0.331)	(0.306)	(0.307)	(0.973)		
WasRP-dummy	0.0460	0.0948	0.0950	1.242		
	(0.333)	(0.331)	(0.333)	(1.032)		
Recession-dummy	0.334***	0.640***	0.630***	0.556***	0.624***	0.548***
	(0.0905)	(0.128)	(0.128)	(0.170)	(0.122)	(0.165)
Rec x CasRP		-0.386***				
		(0.143)				
Rec x CasRP-F		-0.637***				
		(0.195)				
Rec x WasRP		-0.137				
		(0.184)				
Rec x CasRP x Wage cut			-0.548***	-0.407*	-0.547***	-0.402*
			(0.141)	(0.213)	(0.137)	(0.209)
Rec x CasRP x No wage cut			-0.222	-0.148	-0.214	-0.137
			(0.172)	(0.223)	(0.168)	(0.219)
Rec x CasRP-F x Wage cut			-0.798***	-0.623***	-0.789***	-0.606**
			(0.167)	(0.241)	(0.163)	(0.236)
Rec x CasRP-F x Wage cut			-0.519**	-0.468	-0.527**	-0.472
			(0.253)	(0.285)	(0.251)	(0.281)
Rec x WasRP x Wage cut			-0.378**	-0.401*	-0.379**	-0.404*
			(0.156)	(0.210)	(0.154)	(0.211)
Rec x WasRP x Wage cut			0.106	0.177	0.111	0.184
			(0.212)	(0.242)	(0.209)	(0.240)
$Wage^2$	✓	✓	✓	✓	✓	✓
$Wage*CasRP$ & $Wage^2*CasRP$				✓		✓
$Wage*CasRP\text{-}F$ & $Wage^2*CasRP\text{-}F$				✓		✓
$Wage*WasRP$ & $Wage^2*WasRP$				✓		✓
Period & $Period^2$	✓	✓	✓	✓	✓	✓
Constant	-2.967***	-3.091***	-3.047***	-3.590***	-2.821***	-2.779***
	(0.498)	(0.491)	(0.490)	(0.860)	(0.494)	(0.396)
Observations	2858	2858	2858	2858	2858	2858
R^2	0.517	0.518	0.519	0.520	0.627	0.622

Notes: Panel random effects (RE) and fixed effects (FE) regressions on effort for all treatments. Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

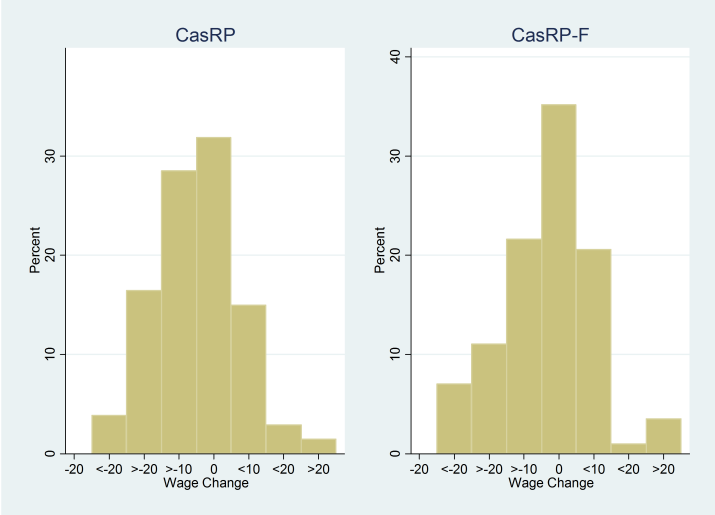


Figure B.1: Histograms of Wage Changes - CasRP vs. CasRP-F

B.2 Hypotheses

In this section, I formally derive the theoretical predictions of the different hypotheses using preferences of Fehr and Schmidt (1999) and the framework of Hart and Moore (2008).

Hypothesis 1

I will formalize the intuition of the first hypothesis using preferences of Fehr and Schmidt (1999) (henceforth FS). The general idea is the following: If workers are sufficiently fair-minded/ inequity-averse, they will reciprocate high wages by high effort, because otherwise inequality would arise lowering workers' utility. In recession, ceteris paribus only the firm's profit decreases. Hence, not cutting wages would lead to higher inequality, (potentially) lowering the worker's utility. So, if workers are sufficiently fair minded, the split of surplus has to be adjusted by cutting wages. Because FS's preferences only focus on outcomes, there are no differences between treatments. All treatments have the same monetary incentives.

In the following, I assume that only workers have FS-preferences. Firms are instead assumed to have standard preferences. This is done because workers' preferences are decisive for the outcome of the gift-exchange game. Firms could pay high wages because they are fair minded or because they are selfish but anticipate the reaction of fair-minded workers correctly. Hence, fair-minded firms would only complicate the analysis without providing much additional insight. I assume that workers are potentially fair-minded and have preferences of the following type:

$$U_W(x_W, x_F) = x_W - \alpha \cdot \max[x_F - x_W, 0] - \beta \cdot \max[x_W - x_F, 0] \tag{B.1}$$

where x_W and x_F are the monetary payoff of the worker (W) and the firm (F) and α reflects how much workers suffer from disadvantageous inequality whereas β reflects how much workers suffer from advantageous inequality. FS additionally assume that $\beta \leq \alpha$ and that $0 \leq \beta < 1$.

In my setting, monetary payoffs of workers and firms are given by $x_W = w - c(e)$, $x_F = \max(10e - w + 50, 0)$ (out of recession) and $x_F^{Rec} = \max(10e - w + 30, 0)$ (in recession). Hence:

$$\begin{aligned} U_W(w, e) \quad = \quad & w - c(e) - \alpha \cdot \max\left[\max(10e - w + 50, 0) - (w - c(e)), 0\right] \\ & -\beta \cdot \max\left[(w - c(e)) - \max(10e - w + 50, 0), 0\right] \end{aligned} \tag{B.2}$$

where the firm is not hit by a recession. Because $\beta \leq \alpha$, workers try to avoid disadvantageous inequality (weakly) stronger than advantageous inequality. Additionally, because the cost of effort is increasing, workers will always lower their effort when their firm earns more than they themselves do. Actually in my setting, workers can hence

always avoid disadvantageous inequality by choosing the minimal effort level (leading to at least the same payoff for the worker as for the firm).

Workers' preferences lead to three different cases: First, if workers are sufficiently fair minded ($\beta > \frac{18}{74} \approx 0.243$), firms offer a high wage and workers reciprocate by a high effort ($w = 84, e = 10$). In recession, the negative profit shock is equally shared and hence the wage is cut by 10 points ($w_{Rec} = 74, e_{Rec} = 10$). Second, if workers are only fair-minded to a very low degree ($\beta < \frac{1}{11} \approx 0.090$), firms offer the minimum wage and workers exert minimal effort ($w = 30, e = 1$). And in recession, wages cannot be cut further. Third, if workers are fair-minded to an intermediate degree ($\frac{1}{11} \leq \beta \leq \frac{18}{74}$), intermediate equilibria can arise in which intermediate wage offers are reciprocated by intermediate effort levels. In recession, wages are then cut by at least 10 points.

The logic behind the first case is the following: Firms offer the highest possible wage that equalizes payoffs and to which workers can react with the highest possible effort ($w = 84, e = 10, x_W = x_F = 66$). Providing a higher wage cannot induce higher effort and providing a lower wage will reduce the worker's effort choice because otherwise disadvantageous inequality results for the worker, leading to lower payoffs for the firm. Workers reciprocate the high wage with the highest effort level if they are sufficiently fair-minded, in this case $\beta > \frac{3}{13} \approx 0.230$. Providing one unit less effort will save the worker (at most) 3 points but advantageous inequality will rise from 0 to 13 points.

In recession, firms again offer the highest wage that equalizes payoffs (accounting for the negative profit shock) and to which workers respond with the highest possible effort ($w^{Rec} = 74, e^{Rec} = 10, x_W^{Rec} = x_F^{Rec} = 56$). Or in other words, the wage is cut by 10 points. Note however, that in recession workers have to be slightly more fair-minded, $\beta > \frac{18}{74} \approx 0.243$. Again, providing one unit less effort will saves the worker 3 points but increases advantageous inequality by 13 points. Importantly however, firms cannot make losses in my setting which restricts the degree of advantageous inequality. Receiving a wage of 74 and providing $e = 1$ instead of $e = 10$ saves the worker 18 points and increases advantageous inequality only by 74 points.[1] Overall, if workers are sufficiently fair-minded ($\beta > \frac{18}{74}$), the results of the first hypothesis arise: firms offer a high wage that equalizes payoffs when reciprocated by high effort and the wage is cut by 10 points in recession.

The logic behind the second case is the following: If workers are only fair-minded to a very limited degree ($\beta < \frac{1}{11} \approx 0.090$), they will always provide minimal effort. Workers can always (at least) save 1 point by lowering effort by one unit which would increase advantageous inequality by 11 points. Hence, firms should only offer the minimum wage, in and out of recession.

For fair-minded workers in between of these thresholds ($\frac{1}{11} \leq \beta \leq \frac{18}{74}$), firms might

[1] Out of recession, receiving a wage of 84 and providing $e = 1$ instead of $e = 10$ saves the worker again 18 but increases advantageous inequality by 84 points. Because $\frac{18}{84} < \frac{3}{13}$, this constraint is, however, not binding.

offer wages lower than $w = 84$ (or $w^{Rec} = 74$) that potentially equalizes payoffs with a lower effort level. This is possible because the cost-of-effort function is increasing. With an effort level of e.g. $e \leq 8$, workers can at most save 2 points (not 3 points) and increase advantageous inequality by 12 points. Hence, reciprocating workers can be less fair-minded compared to my first case. Importantly however, intermediate workers, will always be offered a wage in recession that is at least 10 points lower than the wage out of recession. The logic is the same as for more fair-minded workers: the split of surplus has to be adjusted. Only because workers have to be slightly more fair-minded in recession (firms' losses are cut because negative profits are set to 0) it is also possible that wage cuts are even higher than 10 points because for some levels of workers' β it pays to incentivize workers out of recession but not (or only to a lower degree) in recession.

So overall it holds, that wages are cut by 10 points if workers are sufficiently fair-minded and obviously this wage cut does not differ between treatments because monetary incentives do not change between treatments. Sufficiently fair-minded workers react to this wage cut by not decreasing their effort so that the loss due to a profit shock is equally shared. In the first hypothesis, I do not make such a clear-cut prediction regarding effort. I just predict that controlling for wages, workers provide at least more effort in recession than out of recession but not necessarily sufficiently more effort so that the profit shock is equally shared.

In the experiment, some subjects are potentially fair-minded whereas others are selfish. For this purpose, Kocher and Strasser (2011, App. A1) offer a two-type (one fraction of workers is fair-minded, one fraction is selfish) analysis of their gift exchange setting (see Fehr and Schmidt 1999 and Fehr et al. 2007 for similar models). I do not develop such a model here, but the intuition of these models is quite clear: It is rational both for selfish and fair-minded firms to propose a high wage ($w = 84/\ w_{Rec} = 74$) that equalizes payoffs with high effort ($e = e_{Rec} = 10$) if (fair-minded workers are sufficiently fair-minded and if) the fraction of fair-minded workers is high enough.

Hypothesis 2

The general idea of Hart and Moore (2008) (henceforth HM) is that contracts create feelings of entitlement. Hence, workers that feel entitled to the initial contract conditions in the CasRP treatment might react with shading when firms cut wages. In their paper however, HM do not explicitly deal with renegotiations in their model but focus on comparing different types of contracts. As already outlined by Bartling and Schmidt (2014, p. 18), HM (p. 29-32) verbally consider three different views of renegotiations. Their preferred view assumes that "any flexibility in the trading price must be built into the initial contract." Hence, initially concluded contracts without any flexibility included should serve as strong reference points that cannot be ignored. HM claim that this view is in line with social custom and legal practice. The two other views they consider basically

state that renegotiations undermine the role of contracts as reference points, leading to the effect that in my setting firms should cut wages as suggested by the model of FS. Following Bartling and Schmidt (2014), my hypothesis takes an intermediate position between the outlined views: For the wage rigidity setting, this means that initially concluded contracts (before recessions) may neither lead to complete wage rigidity nor to fully flexible wages. Instead, some workers may react sharply to wage increases, others might not, leading only to more rigid but not completely rigid wages in the CasRP treatment. We might additionally not expect completely rigid wages because firms have an objective reason to cut wages, the recession. This potentially weakens the power of the initial wage as a reference point for the wage in recession.

How can we formalize these ideas? In situations without initial contracts (BASE treatment), I will basically assume that FS-preferences, as outlined for Hypothesis 1, govern the relationship between workers and firms. With this assumption, I deviate form HM who propose that in situations without (initial) contracts, parties feel entitled to the whole surplus or the highest possible payoff. Importantly, HM do not consider the gift-exchange setting. In this setting such an assumption would be highly questionable: Firms would feel entitled to a high effort and a low wage ($w = 30, e = 10, x_F = 120$) whereas workers would feel entitled to a high wage and a low effort ($w = 100, e = 1, x_W = 100$). According to HM, getting less than what you feel entitled to would then lead to shading. Assuming that workers and firms both expect the highest possible payoff for themselves, however, implies that contract parties do not have any shading possibilities in my setting because firms already pay the minimum wage and workers provide the lowest possible effort level. Hence, I do not follow HM's idea in this case.

In the CasRP treatment, contracts can serve as reference points for wage expectations and hence workers' preferences as described in equation (B.1) can be modified in line with HM in the following way:

$$\begin{aligned} U_W(x_W, x_F) &= x_W - \alpha \cdot \max[x_F - x_W, 0] - \beta \cdot \max[x_W - x_F, 0] \\ &\quad -\gamma \cdot I_{[w_2<w_1]} \max[\theta \cdot (w_1 - w_2) - 10 \cdot (e_1 - e_2), 0] \end{aligned} \tag{B.3}$$

where the last term introduces the reference-point effect similar to HM (p. 8) and w_1, w_2 (e_1, e_2) describe wages (effort) out of and in recession (asked for in *stage 1* and *stage 2* of the CasRP treatment). The basic idea of the fourth term is the following: The weighted difference between the reference point wage out of recession and the actual wage in recession, $\theta \cdot (w_1 - w_2)$ can be interpreted as the worker's aggrievement caused by getting less than what you are entitled to and θ represents a weighting parameter. Workers can offset their aggrievement, $\theta \cdot (w_1 - w_2)$, by adjusting their effort downward and hence punishing the firm by $10 \cdot (e_1 - e_2)$. Deviating from HM, I assume that $\theta > 1$, meaning

that workers always have to punish firms more than they lose by a wage cut to offset the negative effects of the reference-point violation. Presumably, $\theta > 1$ is not true for every subject. What is important is that it holds on average. Because assuming this, makes wage cuts non-profitable for firms if workers try to offset their aggrievement. Firms then lose more by reduced effort than they gain by paying a lower wage. Whether workers try to offset their aggrievement finally depends on γ. This parameter determines how much workers weight the negative effect of getting less than what they feel entitled to especially compared to advantageous inequality weighted by β.

What is the optimal strategy for firms if workers have preferences as in (B.3) and workers are sufficiently fair-minded to reciprocate high wages ($\beta > \frac{18}{74}$)? Cutting wages increases the worker's utility loss due the reference-point violation. Not cutting wages, however, also increases the worker's utility loss due to advantageous inequality. Hence, if workers weight the negative impact of violating the reference point high enough ($\gamma \geq 10$), firms should not cut wages in recession. To see why this is the case, assume for the moment that firms and workers act according to the FS-considerations of Hypothesis 1 out of recession ($w = 84, e = 10, x_W = x_F = 66$) and consider the case that firms cut wages in recession by one unit, $w^{Rec} = 83$. The optimal response of workers depends on how much the worker weights the loss of utility due to the reference-point violation. If $\gamma > 10$, even the smallest wage reduction of one unit will lead to a loss of utility of more than 10 if the effort remains unchanged. Reducing the effort by one unit will offset this negative effect of the reference-point violation and lead at most to a loss of utility of less than 10. This is because reducing the effort by one unit increases advantageous inequality by 13 points but the workers still saves 3 points of cost of effort, leading to a net effect smaller than 10 points because β is assumed to be smaller than 1. Hence, if γ is high enough, workers will always decrease their effort if their reference point is violated. However, because I assume $\theta > 1$, a wage reduction is always followed by a sufficiently large reduction of effort that leads to a net reduction in firms' payoffs. Hence, if workers weight the negative impact of violating their reference point to a sufficient degree, firms will not cut wages ($w^{Rec} = 84, e^{Rec} = 10, x_F^{Rec} = 46, x_W^{Rec} = 76$).

If subjects do not care at all about reference points ($\gamma = 0$), firms act according to Hypothesis 1 and cut wages by (at least) 10 points. If workers care to some degree about reference points ($0 < \gamma < 10$), workers trade off the loss of utility that is due to the violation of the reference point and the loss of utility that is due to increased advantageous inequality (caused by a decrease in effort that would offset the first loss of utility). Depending also on the worker's β, it is possible that some workers do not punish small wage cuts by decreasing effort. With a small γ, a small wage cut only leads to a small loss in utility that is due to the reference-point violation. If a worker then has a high β and wants to offset the small utility loss due to the reference-point violation by cutting his effort, this increases the disutility that is due to risen inequality. And if the worker's β

is high enough, cutting effort might hence lead to an overall lower utility of the worker than not cutting effort. In this case, it is optimal for firms to decrease wages slightly. The size of the wage cut then depends on the worker's γ and β or the firm's expectation of these values.

So overall, depending on how strong reference points are for workers, firms should not cut wages, cut them only slightly or fully adjust wages downward. In Hypothesis 2, I suggest an intermediate case, that wages in the CasRP treatment should at least be more rigid than in the BASE treatment. Regarding the effort, the model with workers sufficiently caring about reference-point violations provides a clear cut prediction: In equilibrium, firms do not adjust their wage and workers do not adjust their effort. Hence, controlling for the wage level, workers should exert the same effort in and out of recession in equilibrium. Importantly however, in equilibrium, effort is not increased in recession although wages remain stable because workers already exert the maximum level of effort. Exerting even a higher effort (if possible) would further decrease advantageous inequality. I try to modify this clear-cut prediction for Hypothesis 2 by incorporating some out-of-equilibria behavior. Firms may e.g. need some time to learn how workers react to wage changes. The model predicts that workers decrease their effort if wages are cut too strong (and workers care sufficiently about the reference-point violation). So assuming that not all subjects play the equilibrium strategy, the crucial prediction regarding effort is the following: In the CasRP, a reference-point effect should be observed. This means that controlling for the wage level wage cuts should be accompanied with relatively lower effort compared to stable (or increased wages).

An important aspect that has been neglected so far is that firms could anticipate that a high wage out of recession is problematic because it lowers payoffs in recession. They could react to this by paying lower wages (and hence foregoing earnings) out of recession and hence being able to pay lower wages (and hence gain earnings) also in recession. This kind of strategy is, however, suboptimal given the parameters of the experiment. Paying $w = 84$ out of and in recession leads to earnings of $x_F = 66$ and $x_F^{Rec} = 46$ with a probability of 2/3 and 1/3 respectively. The best strategies available, in which the out of recession payment is reduced, lead to payoffs below this result. Paying $w = 78$ out of and in recession (in order to (almost) equalize payoffs with an effort level of 9) leads to earnings of $x_F = 62$ and $x_F^{Rec} = 52$: A loss of 4 out of recession with a probability of 2/3 and a gain of 6 in recession with a probability of 1/3. Paying $w = 71$ out of recession (in order to equalize payoffs with an effort level of 8) and paying $w = 74$ in recession leads to payoffs of $x_F = 59$ and $x_F^{Rec} = 56$: A loss of 7 out of recession with probability 2/3 and a gain of 10 in recession with probability of 1/3. Again, the low likelihood of recession makes this strategy suboptimal. Crucially, however, these strategies may become optimal for firms that also have preferences of the FS-type. In equilibrium, not cutting wages leads to disadvantageous inequality, lowering firms's utility compared to standard preferences.

Lowering the wage out of equilibrium leads to less disadvantageous inequality in recession. Without making specific assumptions on the parameters for the inequality averse firms, it is difficult to derive a clear prediction whether they should lower wages out of recession compared to the BASE treatment, but it remains a possibility. Hence, I consider it as an open question.

Finally, following HM, I have assumed so far that workers feel basically fully entitled to the status-quo wage. Assuming, however, that workers feel entitled to their rational expectations following Köszegi and Rabin (2006) would complicate the analysis quite a lot but not change predictions substantially because even when expectations serve as reference points, wage cuts are still perceived as reference point violations. With expectations determining the reference point, workers' reference points would incorporate both the recession wage and the out-of-recession wage with the appropriate probability. Importantly, if the recession wage is adjusted downward, workers still compare this adjusted wage partially with the out-of-recession wage. Hence, wage cuts would still be considered as a reference-point violation and if workers care enough for their reference point in expectations, they will punish these reference point violations.

Hypothesis 3

A crucial aspect not considered so far is that in the analysis above, firms have correct beliefs about workers' behavior and their parameters. Of course, this might not be the case especially at the beginning of the experiment. Firms learn about workers' behavior during the course of the experiment. Due to the information restrictions in the CasRP treatment, this learning might not be sufficient to lead to optimal behavior. If firms e.g. underestimate how much workers care about reference-point violations (if they underestimate γ in equation B.3), they might adjust wages downward as in the BASE treatment. Receiving more information in CasRP-F might facilitate learning about the γ-parameter and hence lead to more wage rigidity.

Importantly however, not only firms' behavior might change in the CasRP-F treatment, also workers might behave differently because they now have a punishment device that is easily interpretable by firms. Hence, workers might use this device more intensively. In the model, such an effect would potentially be captured by a higher θ. With a higher θ, offsetting the aggrievement caused by a wage cut needs a stronger reduction in effort than with a lower θ, making wage cuts even less profitable.

B.3 Additional Analysis: Payoffs

This section provides an additional analysis regarding subjects' payoffs. First, I will consider firms' payoffs. Table B.4 provides average payoffs of firms for different treatments, in and out of recession. Additionally, for all treatments besides the BASE treatment, recession profits are separated for wage cuts versus stable (or increased) wages. The last two columns provide two-sided non-parametric Wilcoxon signed rank tests that analyze (on the individual and on the session level) whether there is a difference in payoffs when firms cut wages or when they pay stable (or increased) wages.[2]

Table B.4: Average Firm Payoffs: All Treatments

	No-Recession	Recession		Wage cut vs. no wage cut	
		Wage cut	No wage cut	p-values (Wilcoxon)	
				Individual level	Session level
BASE	31.6		17.8		
CasRP	33.8	16.6	18.4	0.058	0.463
CasRP-F	33.5	11.7	19.5	0.239	0.028
WasRP	32.4	15.4	21.6	0.055	0.035

Notes: The last two columns report two-sided p-values of a Wilcoxon signed rank test that evaluates whether the payoff distribution is different between wage cuts and stable (or increased) wages.

Compared to the BASE treatment, out-of-recession payoffs are only slightly higher for firms in the CasRP, the CasRP-F, and the WasRP treatment, and this difference is only significant for the CasRP treatment when testing on the individual level.[3] Hence, overall, payoffs out-of-recession are fairly similar between treatments. Looking now at the recession, one can see that payoffs are in general lower when wages are cut than when they are not cut and this difference is (at least marginally) significant in 4 out of 6 cases. Surprisingly, the difference between wage cuts and stable (or increased) wages is highest for the CasRP-F treatment. On first glance, this seems at odds with the finding that I was unable to detect a statistically significant reference-point effect in workers' behavior. In the CasRP-F treatment, workers do not reward stable (or increased) wages compared to wage cuts. Still, those firms that cut wages in the CasRP-F treatment make much lower profits than those firms that do not cut wages. A potential explanation lies in the fact that average wages in recession in the CasRP-F treatment are much smaller for wage cuts than for stable (or increased) wages (Wage cuts: $w^{Rec} = 44.7$, stable or increased wages: $w^{Rec} = 59.1$).[4]

[2] In principle, the CasRP Hypothesis provides a directed prediction, namely that firms that cut wages should earn less than those firms that pay stable (or increased) wages. Nonetheless, I report two-sided p-values for matters of consistency because the directed hypothesis seems not to hold for workers, as will be outlined below.

[3] Two-sided Wilcoxon signed rank test: BASE vs. CasRP - individual: $p = 0.018$, session: $p = 0.200$; BASE vs. CasRP-F - individual: $p = 0.149$, session: $p = 0.422$; BASE vs. WasRP - individual: $p = 0.347$, session: $p = 0.873$.

[4] Similar difference are also present in the CasRP and the WasRP treatment (CasRP - Wage cuts: $w^{Rec} = 48.4$, stable or increased wages: $w^{Rec} = 55.3$; WasRP - wage cuts: $w^{Rec} = 46.9$, stable or increased

Table B.5: Panel Regression on Firms' Profits

	(1) CasRP	(2) CasRP	(3) CasRP-F	(4) CasRP-F	(5) WasRP	(6) WasRP
Wage	0.114		-0.171		0.408	
	(0.535)		(0.433)		(0.214)	
Wage cut	-4.497*	-6.698**	-2.586	-2.400	-4.925*	-3.321
	(2.139)	(2.506)	(2.998)	(3.317)	(1.892)	(2.215)
$Wage^2$	✓		✓		✓	
Wage-dummies		✓		✓		✓
Period & $Period^2$	✓	✓	✓	✓	✓	✓
Constant	9.910	18.97	21.50*	24.69**	8.619	22.30**
	(14.95)	(11.09)	(9.996)	(8.873)	(6.301)	(6.655)
Observations	207	207	199	199	216	216
R^2	0.053	0.106	0.128	0.187	0.112	0.136

Notes: Panel fixed effects (FE) regressions on firms' profits for the CasRP, the CasRP-F, and the WasRP treatment (only for recession data). Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

To control for the wage level when comparing wage cuts with stable (or increased) wages, Table B.5 provides panel (fixed effects) regressions on firms' payoffs. All regressions only use recession data. Regressions (1)-(2) only use the data for the CasRP treatment, (3)-(4) only use the data for the CasRP-F treatment, and (5)-(6) only use the data for the WasRP treatment. Focusing only on recession data separately for every treatment reduces the number of observations sharply but it allows for controlling for the wage-effort relationship in recession to the best extent.[5] For each treatment, two different regressions are provided. First, one regression just controls for the wage and a quadratic wage term. Second, one regression uses the wage dummies already used in Table 3.5. Effort levels are not included because combined with the wage they would determine payoffs. In all regressions, the wage coefficients are insignificant, potentially because they are imprecisely measured due to the fairly low number of observations in recession.

Regressions (1)-(2) show that in the CasRP treatment wage cuts lead to lower profits in recession than stable (or increased) wages. This difference is at least marginally significant for both regression specifications. Regression (3)-(4) show that the difference between wage cuts and stable (or increased) wages is not significant in the CasRP-F treatment. So, if one controls carefully for the wage level in recession, there does not seem to be a statistically significant difference in payoffs between stable (or increased) wages and wage cuts although firms cutting wages earn on average much less profit in this treatment. But these firms also pay much lower wages. This result corresponds to the fact that I do not find a statistically significant reference-point effect in the CasRP-F treatment as in the CasRP treatment. Importantly however, the wage cut dummy at least has the expected sign. Profits fall if wages are cut. Finally, regressions (5)-(6) show that wage cuts lead to

wages: $w^{Rec} = 60.6$). These difference are, however, smaller than in the CasRP-F treatment.

[5] Using not only recession but also out-of-recession data leads to slightly different results. In this case, wage-cut dummies for the CasRP-F and the WasRP treatment are significant but not for the CasRP treatment. Nonetheless, I believe that using only recession data is most appropriate in this case because only then the actual wage-effort relationship is captured in the best way.

Table B.6: Average Worker Payoffs: All Treatments

	No-Recession	Recession		Wage cut vs. now wage cut	
		Wage cut	No wage cut	p-values (Wilcoxon)	
				Individual level	Session level
BASE	55.1		43.3		
CasRP	52.8	44.9	50.0	0.099	0.463
CasRP-F	53.8	43.8	53.0	0.001	0.028
WasRP	55.8	43.8	53.9	0.001	0.028

Notes: The last two columns report two-sided p-values of a Wilcoxon signed rank test that evaluates whether the payoff distribution is different between wage cuts and stable (or increased) wages.

lower payoffs compared to stable or increased wages in the WasRP treatment. This effect is, however, only statistically significant for one regression specification. This result is in line with the finding that there is a reference-point effect in the WasRP treatment. So overall, the regression analysis confirms the findings of the main text: In the CasRP and the WasRP treatment, wage cuts lead to lower payoffs compared to stable (or increased) wages (if one carefully controls for the wage level). For the CasRP-F treatment, the wage cut dummy has the expected sign but is not significant.

A similar analysis can be done for workers. Table B.6 provides the same data for workers that Table B.4 provides for firms. Out of recession, profits are again fairly similar and this time no difference is significant.[6] In recession, we observe the same pattern in all treatments, in which a reference-point effect can occur: Workers' profits are lower when the wage has been cut compared to stable (or increased) wages, and these differences are significant at least for the CasRP-F and the WasRP treatment. In principle, I expect that wage cuts lead to lower effort, increasing workers' payoff. Importantly, however, wages that are cut are on average lower than wages that are not cut. Hence, workers on average earn more if wages are not cut, even if they exert more costly effort in this case.

Table B.7 provides regressions in a similar way than those of Table B.5. This time, however, workers' profits are considered and regressed on wages. In each regression, I expect a positive sign for the *wage-cut* dummy. The coefficient of the dummy should, however, be fairly small. In many cases workers observe only a moderate wage cut and hence cut the effort often only by one point. A one point effort reduction, however, only increases workers' payoffs by one or up to three points depending on how much effort the worker had chosen before. Regressions (1)-(2) show that wage cuts in the CasRP treatment lead to higher worker payoffs although the effect is not significant. The same is true for CasRP-F treatment (regressions (3) and (4)). Only for the WasRP treatment, workers' payoffs significantly increase when wages are cut, as can be seen in regressions (5)-(6). These results suggest that controlling for the wage level, workers seem to profit from wage cuts, because they decrease their effort, but the effect is very small (and only

[6]Two-sided Wilcoxon signed rank test: BASE vs. CasRP - individual: $p = 0.140$, session: $p = 0.200$; BASE vs. CasRP-F - individual: $p = 0.257$, session: $p = 0.423$; BASE vs. WasRP - individual: $p = 0.764$, session: $p = 0.873$.

Table B.7: Panel Regression on Workers' Profit

	(1) CasRP	(2) CasRP	(3) CasRP-F	(4) CasRP-F	(5) WasRP	(6) WasRP
Wage	0.975***		1.013***		0.797***	
	(0.130)		(0.069)		(0.050)	
Wage cut	0.567	0.491	0.391	0.209	1.192***	1.208***
	(0.322)	(0.327)	(0.540)	(0.431)	(0.223)	(0.120)
Wage^2	✓		✓		✓	
Wage-dummies		✓		✓		✓
Period & Period^2	✓	✓	✓	✓	✓	✓
Constant	2.075	85.25***	1.147	84.07***	6.048**	88.40***
	(3.611)	(3.086)	(2.110)	(1.796)	(1.689)	(2.813)
Observations	207	207	199	199	216	216
R^2	0.953	0.952	0.944	0.948	0.946	0.954

Notes: Panel fixed effects (FE) regressions on workers' profits for the CasRP, the CasRP-F, and the WasRP treatment (only for recession data). Standard errors are reported in parentheses, adjusted for clustering at the session level. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

sometimes significant), as expected.

Overall, the presented results in this section confirm the findings of the main text. Although differences in payoffs are not overwhelmingly large and not always significant, there is some evidence that cutting wages leads to lower payoffs for firms and slightly higher payoffs for workers when controlling for the wage level. Hence, differences in payoffs are fairly small but qualitatively in line with a reference-point effect.

B.4 Instructions

This section provides the translated instructions for all treatments. The general structure/ design of the instructions is based on Charness et al. (2012) (and partially also on Gerhards and Heinz (2012)), although obviously large modifications had to be made to capture the specifics of my design. In the English version of these instructions, I will consider for convenience the employer as female and the employee as male. The instructions for the BASE and the CasRP treatment mainly differ with respect to describing the sequence of events. Additionally, the description of the payoffs and the description of the examples have to be slightly modified in the CasRP treatment. The instructions of the CasRP-F treatment are very similar to those of the CasRP treatment. Only the description of *Stage 3* is different, as can be seen below. The instructions for the WasRP treatment are a mixture of the instructions of the BASE and the CasRP treatment.

Baseline (BASE) Instructions

Welcome to our experiment! Please read the following instructions carefully. Your final payoffs will also depend on how good you have understood these instructions. You can ask questions anytime, just raise your hand. But please do not speak with other participants of the experiment any longer.

Overview

You participate in a study about the labor market. One half of all participants will be in the role of an **employer**, and one half of all participants will be in the role of an **employee**. Which role you take, will be decided randomly and you are informed about your role at the beginning of the actual experiment. You will keep your role until the end of the experiment.
The **general structure** of the experiment is the following: The employer offers a **wage** to her employee. The employee then decides whether he would like to accept this offer. If he accepts, he then has to decide how much **effort** he would like to exert. The higher his effort, the more his company benefits.
Another important aspect is that the employers' earnings are not constant throughout the experiment. Every period, a **bad market situation** can occur for an **employer** with a probability of 1/3. Then, the employer's earnings are lower than before.

The Details

The experiment consists out of **18** periods. In every period, you will be randomly matched with another participant of the experiment. The matching works such that you will never meet the same participant in two consecutive periods. Additionally, **one employer is always exactly matched with one employee** and no participant will ever learn the identity of his or her counterpart.

All of the 18 periods proceed in the following way:

Stage 1: (Contract Offer)
At the beginning of the first stage, the computer **randomly** chooses two out of six employers. Only for these two employers a **bad market situation** occurs, e.g. the earnings of these employers are reduced by **20** points. For all other employers **no** bad market situation occurs. Independent of whether a bad market situation has occurred or not, every **employer** then has to make a **contract offer** to her matched employee by choosing as a **wage** an integer between **30** and **100**.

Stage 2: (Contract Conclusion and Effort)
First of all, the offered wage of his employer is now communicated to the employee. Afterwards, two cases have to be distinguished:

1. **No** bad market situation (for the employer): In this case, the **employee** initially has to decide whether he wants to accept the contract offered by his employer. If he declines the offer, both employer and employee learn their earnings in *Stage 3*, and for them this period is over. If he on the contrary accepts the offer, a **contract** is concluded, and the employee has to choose how much effort he would like to exert. Here, **effort** is a number between **1** and **10**. The higher the effort is, the higher the employee's costs are (compare the Table):

Effort	1	2	3	4	5	6	7	8	9	10
Costs	0	1	2	4	6	8	10	12	15	18

2. **Bad market situation** (for the employer): In this case, the **employee** has to **accept** the contract offered by the employer and hence a contract is always concluded. The employee only has to choose how much effort he would like to exert. Just like in 1., **effort** is a number between **1** and **10**, and the costs of effort are the same as in the Table of 1.

Stage 3: (Earnings)
Now, employees and employers are informed about their (period) earnings. If a contract has been concluded (*Stage 2*), the employer also learns her employee's effort level.

Payoffs

If a contract has not been concluded, the **employer** gets no earnings (= 0 points) in this period. If a contract has been concluded, the following holds: Wage and effort determine the **employer's** earnings. In the case that a **bad market situation** has occurred, the employer's earnings are moreover **20 points** lower than in the case that no bad market situation has occurred:

Employer (no bad market situation) = (10 * effort) - wage + 50

Employer (bad market situation) = (10 * effort) - wage + 30

In both cases we have: The higher the employee's effort level and the lower the wage, the higher the employer's earnings are. Additionally, in both cases it holds that, if the formula leads to negative earnings, the employer gets **0** points. Hence, she can not make losses.

If a contract has not been concluded, the **employee**, contrary to the employer, gets positive earnings of 20 points.
If a contract has been concluded, the following holds: Wage and the costs of effort determine the **employee's** earnings. These earnings are not different between both market situations:

Employee (no bad market situation) = Wage - costs of effort
Employee (bad market situation) = Wage - costs of effort
In both case we have: The higher the wage and the lower the effort (or rather the lower the costs of effort), the higher the employee's earnings are.

At the end of the experiment, you will be be paid individually and confidentially. Your payoff will be the sum of all earnings gained in all periods. All points earned in the different periods will be converted in the following way: **35 points = 100 Eurocent**.

Four Examples

1. Situation: A bad market situation does **not** occur! Employer's contract offer: Wage = 50; Employee declines offer!
Employer's earnings = 0 points
Employee's earnings = 20 points

2. Situation: A bad market situation does **not** occur! Employer's contract offer: Wage = **30**; Employee accepts offer and exerts effort= **1**
Employer's earnings = 10 * **1** - **30** + 50 = 30
Employee's earnings = **30** - 0 = 30

3. Situation: A bad market situation does **not** occur! Employer's contract offer: Wage = **50**; Employee accepts offer and exerts effort= **5**
Employer's earnings = 10 * **5** - **50** + 50 = 50
Employee's earnings = **50** - 6 = 44

4. Situation: A bad market situation does occur! Employer: Wage = **60**; Employee: Effort = **7**
Employer's earnings = 10 * **7** - **60** + 30 = 40

Employee's earnings = **60** - 10 = 50

Further course of events

When you have finished reading the instructions, please first try to deal with the additional example situations provided at your computer screen. Of course, you can ask questions at any time. When all participants have solved these situations correctly, three **unpaid** trial rounds follow in order to make you familiar with the different computer screens. For reasons of simplicity, the probability of a bad market situation is set to zero in these trial rounds. Afterwards, the 18 periods of the actual experiment start.
A last hint: Because a bad market situation really occurs randomly for employers, a single employer can be hit by such a bad market situation also in consecutive periods and then he may not be hit by such a situation for a longer period of time. Because every employer is exactly matched with one employee, this reasoning also holds for employees. Neither employees nor employers however know how often their counter-part has been hit by a bad market situation in the past.

Contracts as Reference Points (CasRP) Instructions

Welcome to our experiment! Please read the following instructions carefully. Your final payoffs will also depend on how good you have understood these instructions. You can ask questions anytime, just raise your hand. But please do not speak with other participants of the experiment any longer.

Overview

You participate in a study about the labor market. One half of all participants will be in the role of an **employer**, and one half of all participants will be in the role of an **employee**. Which role you take, will be decided randomly and you are informed about your role at the beginning of the actual experiment. You will keep your role until the end of the experiment.
The **general structure** of the experiment is the following: The employer offers a **wage** to her employee. The employee then decides whether he would like to accept this offer. If he accepts, he then has to decide how much **effort** he would like to exert. The higher his effort, the more his company benefits.
Another important aspect is that the employers' earnings are not constant throughout the experiment. Every period, a **bad market situation** can occur for an **employer** with a probability of 1/3. Then, the employer's earnings are lower than before.

The Details

The experiment consists out of **18** periods. In every period, you will be randomly matched with another participant of the experiment. The matching works such that you will never meet the same participant in two consecutive periods. Additionally, **one employer is always exactly**

matched with one employee and no participant will ever learn the identity of his or her counterpart.

All of the 18 periods proceed in the following way:

Stage 1: (Contract Conclusion)

1. The **employer** makes a **contract offer** to her matched employee by choosing as a **wage1** an integer between **30** and **100**.

2. The **employee** initially has to decide whether he wants to accept this contract offer. If he declines the offer, both employer and employee learn their earnings in *Stage 3* and for them this period is over. If he on the contrary accepts the offer, a **contract** is concluded, and the employee has to choose how much effort he would like to exert (whereof the employer is initially not informed about). Here, **effort1** is a number between **1** and **10**. The higher the effort is, the higher the employee's costs are (compare the Table):

Effort	1	2	3	4	5	6	7	8	9	10
Costs	0	1	2	4	6	8	10	12	15	18

   Importantly, a concluded contract is binding for the employee to the extent that he cannot withdraw form the contract in *Stage 2*, even if the employer changes the wage in *Stage 2*

Stage 2: (Potential contract changes)
At the beginning of the second stage, the computer **randomly** chooses two out of six employers. Only for these two employers a **bad market situation** occurs, e.g. the earnings of these employers are reduced by **20** points. For all other employers **no** bad market situation occurs.

1. If a bad market situation occurs for the **employer**, she has the opportunity to change the existing contract, e.g. she can change her employee's wage in any way. For this purpose, she choose as a **wage2** again an integer between **30** and **100**. Wage2 and wage1 can be totally different, but also the same values can be chosen for wage1 and wage2.

2. Because an existing contract is binding for him, the employee cannot withdraw from this contract and has to accept any wage change. But also the employee can change his effort in any way. For this purpose, he chooses as **effort2** a number between **1** and **10** (costs are the same as in *Stage 1*). Effort2 and effort1 can be totally different, but also the same values can be chosen for effort1 and effort2.

Employers, for which no bad market situation occurs (and their matched employees) cannot change the contract concluded in *Stage 1* and do not have to provide any input in *Stage 2* and learn their earnings in *Stage 3*.

Stage 3: (Earnings)
Now, employees and employers are informed about their (period) earnings. If a contract has been concluded (*Stage 1*), the employer also learns her employee's **relevant** effort level: Either effort2 in the case of a bad market situation or else effort1.

Payoffs

If a contract has not been concluded, the **employer** gets no earnings (= 0 points) in this period. If a contract has been concluded, the following holds: In the case that no bad market situation has occurred wage1 and effort1 (initial contract conclusion) determine the **employer's** earnings. In the case that a bad market situation has occured, wage2 and effort2 (potential contract change) determine her earnings and these earnings are moreover **20 points** lower:

Employer (no bad market situation) = (10 * effort1) - wage1 + 50
Employer (bad market situation) = (10 * effort2) - wage2 + 30

In both cases we have: The higher the employee's effort level and the lower the wage, the higher the employer's earnings are. Additionally, in both cases it holds that, if the formula leads to negative earnings, the employer gets **0** points. Hence, she can not make losses.

If a contract has not been concluded, the **employee**, contrary to the employer, gets positive earnings of 20 points.
If a contract has been concluded, the following holds: In the case that no bad market situation has occurred wage1 and the costs of effort1 (initial contract conclusion) determine the **employee's** earnings. In the case that a bad market situation has occurred, wage2 and the costs of effort2 (potential contract change) determine his earnings:

Employee (no bad market situation) = Wage1 - costs of effort1
Employee (bad market situation) = Wage2 - costs of effort2
In both case we have: The higher the wage and the lower the effort (or rather the lower the costs of effort), the higher the employee's earnings are.

At the end of the experiment, you will be be paid individually and confidentially. Your payoff will be the sum of all earnings gained in all periods. All points earned in the different periods will be converted in the following way: **35 points = 100 Eurocent**.

Four Examples

1. **Situation:** Employer's contract offer: Wage1 = 50; Employee declines offer!
   Employer's earnings = 0 points

Employee's earnings = 20 points

2. Situation: Employer's contract offer: Wage1 = **30**; Employee accepts offer and exerts effort1 = **1**; A bad market situation does **not** occur (hence no contract changes)
Employer's earnings = 10 * **1** - **30** + 50 = 30
Employee's earnings = **30** - 0 = 30

3. Situation: Employer's contract offer: Wage1 = **50**; Employee accepts offer and exerts effort1 = **5**; A bad market situation does **not** occur (hence no contract changes)
Employer's earnings = 10 * **5** - **50** + 50 = 50
Employee's earnings = **50** - 6 = 44

4. Situation: Employer's contract offer: Wage1 = **50**; Employee accepts offer and exerts effort1 = **5**; A **bad market situation** occurs (Wage and effort are changed); Contract changes: Wage2 = **60**; Effort2=**7**
Employer's earnings = 10 * **7** - **60** + 30 = 40
Employee's earnings = **60** - 10 = 50

Further course of events

When you have finished reading the instructions, please first try to deal with the additional example situations provided at your computer screen. Of course, you can ask questions at any time. When all participants have solved these situations correctly, three **unpaid** trial rounds follow in order to make you familiar with the different computer screens. For reasons of simplicity, the probability of a bad market situation is set to zero in these trial rounds. Afterwards, the 18 periods of the actual experiment start.
A last hint: Because a bad market situation really occurs randomly for employers, a single employer can be hit by such a bad market situation also in consecutive periods and then he may not be hit by such a situation for a longer period of time. Because every employer is exactly matched with one employee, this reasoning also holds for employees. Neither employees nor employers however know how often their counter-part has been hit by a bad market situation in the past.

Contract as Reference Points with Feedback (CasRP-F) Instructions

The instructions for the CasRP-F treatment are very similar to the instructions of the CasRP treatment. They only differ in the text of Stage 3:

Stage 3: (Earnings)
Now, employees and employers are informed about their (period) earnings. If a contract has been concluded (*Stage 1*), the employer only now learns her employee's effort level: In case of a

bad market situation, she learns **both** effort1 **and** effort2 such that the employer can see how her employee **has reacted** to a (potential) wage change. In case that no bad market condition has occurred, the employer only learns effort1.

Wages as Reference Points (WasRP) Instructions

Welcome to our experiment! Please read the following instructions carefully. Your final payoffs will also depend on how good you have understood these instructions. You can ask questions anytime, just raise your hand. But please do not speak with other participants of the experiment any longer.

Overview

You participate in a study about the labor market. One half of all participants will be in the role of an **employer**, and one half of all participants will be in the role of an **employee**. Which role you take, will be decided randomly and you are informed about your role at the beginning of the actual experiment. You will keep your role until the end of the experiment.
The **general structure** of the experiment is the following: The employer offers a **wage** to her employee. The employee then decides whether he would like to accept this offer. If he accepts, he then has to decide how much **effort** he would like to exert. The higher his effort, the more his company benefits.
Another important aspect is that the employers' earnings are not constant throughout the experiment. Every period, a **bad market situation** can occur for an **employer** with a probability of 1/3. Then, the employer's earnings are lower than before.

The Details

The experiment consists out of **18** periods. In every period, you will be randomly matched with another participant of the experiment. The matching works such that you will never meet the same participant in two consecutive periods. Additionally, **one employer is always exactly matched with one employee** and no participant will ever learn the identity of his or her counterpart.

All of the 18 periods proceed in the following way:

Stage 1: (Contract Offer)

1. First of all, the **employer** makes a **contract offer** to her matched employee for the case that no bad market situation occurs by choosing as a **wage1** an integer between **30** and **100**.

2. Afterwards, the computer **randomly** chooses two out of six employers. Only for these two employers a **bad market situation** occurs, e.g. the earnings of these employers are reduced by **20 points**. For all other employers **no** bad market situation occurs.

3. If a bad market condition occurs for the **employer**, she has the opportunity to change the contract offer made in 1. in any way. For this purpose, she chooses as a **wage2** again an integer between **30** and **100**. Wage2 and wage1 can be totally different, but also the same values can be chosen for wage1 and wage2.
   Employers, for which no bad market situation occurs cannot change the contract offer and do not have to provide any further input.

Stage 2: (Contract Conclusion and Effort)
For the **employee** two cases have to be distinguished:

1. **No** bad market situation (for the employer): In this case, the **employee** is informed about wage1 and has initially to decide whether he wants to accept the contract offered by his employer. If he declines the offer, both employer and employee learn their earnings in *Stage 3* and for them this period is over. If he on the contrary accepts the offer, a **contract** is concluded, and the employee has to choose how much effort he would like to exert. Here, **effort1** (no bad market condition) is a number between **1** and **10**. The higher the effort is, the higher the employee's costs are (compare the Table):

Effort	1	2	3	4	5	6	7	8	9	10
Costs	0	1	2	4	6	8	10	12	15	18

2. **Bad market situation** (for the employer): In this case, the **employee** is informed about both wage1 and wage2, where only wage2 is relevant for the payoff. Additionally, he has to **accept** the contract offered by the employer and hence a contract is always concluded. The employee only has to choose, how much effort he would like to exert. Just like in 1., **effort2** (bad market condition) is a number between **1** and **10** and the costs of effort are the same as in the Table of 1.

Stage 3: (Earnings)
Now, employees and employers are informed about their (period) earnings. If a contract has been concluded (*Stage 2*), the employer also learns her employee's effort level.

Payoffs

If a contract has not been concluded, the **employer** gets no earnings (= 0 points) in this period. If a contract has been concluded, the following holds: In case no bad market situation has occurred, wage1 and effort1 (no bad market situation) determine the **employer's** earnings. In the case that a **bad market situation** has occurred, wage2 and effort2 (bad market condition) determine her earnings and these earnings are moreover **20 points** lower:

Employer (no bad market situation) = (10 * effort1) - wage1 + 50
Employer (bad market situation) = (10 * effort2) - wage2 + 30

In both cases we have: The higher the employee's effort level and the lower the wage, the higher the employer's earnings are. Additionally, in both cases it holds that, if the formula leads to negative earnings, the employer gets **0** points. Hence, she can not make losses.

If a contract has not been concluded, the **employee**, contrary to the employer, gets positive earnings of 20 points.
If a contract has been concluded, the following holds: In the case that no bad market situation has occurred wage1 and the cost of effort1 (no bad market situation) determine the **employee's** earnings. In case that a bad market situation has occurred, wage2 and the costs of effort2 (bad market situation) determine the earnings:

Employee (no bad market situation) = Wage1 - costs of effort1
Employee (bad market situation) = Wage2 - costs of effort2
In both case we have: The higher the wage and the lower the effort (or rather the lower the costs of effort), the higher the employee's earnings are.

At the end of the experiment, you will be be paid individually and confidentially. Your payoff will be the sum of all earnings gained in all periods. All points earned in the different periods will be converted in the following way: **35 points = 100 Eurocent**.

Four Examples

1. Situation: Employer's contract offer: Wage1 = 50; A bad market situation does **not** occur! Employee declines offer!
Employer's earnings = 0 points
Employee's earnings = 20 points

2. Situation: Employer's contract offer: Wage1 = **30**; A bad market situation does **not** occur! Employee accepts offer and exerts effort1 = **1**
Employer's earnings = 10 * **1** - **30** + 50 = 30
Employee's earnings = **30** - 0 = 30

3. Situation: Employer's contract offer: Wage1 = **50**; A bad market situation does **not** occur! Employee accepts offer and exerts effort1 = **5**
Employer's earnings = 10 * **5** - **50** + 50 = 50
Employee's earnings = **50** - 6 = 44

4. Situation: Employer's contract offer: Wage1 = **50**; A **bad market situation** occurs
Contract offer is changed: Employer: Wage2 = **60**; Employee: Effort2 = **7**

Employer's earnings = 10 * **7** - **60** + 30 = 40

Employee's earnings = **60** - 10 = 50

Further course of events

When you have finished reading the instructions, please first try to deal with the additional example situations provided at your computer screen. Of course, you can ask questions at any time. When all participants have solved these situations correctly, three **unpaid** trial rounds follow in order to make you familiar with the different computer screens. For reasons of simplicity, the probability of a bad market situation is set to zero in these trial rounds. Afterwards, the 18 periods of the actual experiment start.

A last hint: Because a bad market situation really occurs randomly for employers, a single employer can be hit by such a bad market situation also in consecutive periods and then he may not be hit by such a situation for a longer period of time. Because every employer is exactly matched with one employee, this reasoning also holds for employees. Neither employees nor employers however know how often their counter-part has been hit by a bad market situation in the past.

Appendix C

Appendix to Chapter 4

In this appendix, I first provide additional considerations regarding equilibrium play at the boundaries of the signal space (C.1). Then, two additional figures are given (Appendix C.2). Afterwards, I present the translated instructions for the *AuctionFirst* treatment (Appendix C.3). In general, instructions were based on those of Kagel and Levin (1986), although large modifications had to be made to capture our specific experimental design. Original instructions were written in German. Instructions for the *TransformedFirst* treatment are very similar. Additionally, *Frequently Asked Questions* that were orally presented to subjects after explaining the auction game (Part I) and after explaining the transformed game (Part II) are outlined after the Instructions (Appendix C.4).

C.1 Equilibrium at the boundaries

In the main text, equilibrium considerations regarding the auction game do not take into account that subjects receiving a signal close to 25 or 225 can infer the commodity's real value. This might not only influence those subjects' strategies that receive signals close to 25 or 225, but it could in principle also influence those subjects' strategies that receive signals well within the interval. In the following, we will, however, outline why this influence vanishes very quickly and why $b_i = -8$ remains the equilibrium strategy for all realizations of the commodity's value that occur in the experiment. We start with the lower boundary: In order to analyze how subjects' strategy at the boundary influence subjects' strategy for central-value signals, we consider five player types. Player 5 receives a signal $x^5 \in [46, 54)$. His strategy might be influenced by his potential opponent with the lower signal: player 4, who receives the signal $x^4 = x^5 - 6, x^4 \in [40, 46)$. But player 4's strategy might of course be influenced by player 3 ($x^3 = x^4 - 6, x^3 \in [34, 40)$) whose strategy might be influenced by player 2 ($x^2 = x^3 - 6, x^2 \in [28, 34)$) and finally also by player 1 ($x^1 = x^2 - 6, x^1 \in [22, 28)$).

Player 1 receives a signal $x^1 \in [22, 28)$ from which he can infer that the commodity's real value is above his own signal. For this reason, player 1 cannot make any profits from

underbidding by -8. Instead player 1 tries to overbid[1] player 2. But importantly, player 1 bids at most $b^1 = +3$ because otherwise he would lose money because of overbidding the commodity's value $x^1 + 3$. Hence, in equilibrium, player 2 will bid $b^2 \geq -3.01$ because any bid below would provide player 1 with an overbidding incentive that would lead player 2 to adjust his bid upwards. Additionally, player 2 cannot bid more than $b^2 = 0$ because higher bids would lead to negative expected payoffs. Because of these incentives of player 2, in equilibrium, player 3 can ensure himself an expected payoff of at least $Eu_i = 1.495$ by bidding $b^3 = -5.99$. If player 3 follows this strategy, player 2 cannot gain money by winning the auction, and, hence, player 2 will not overbid the player 3 and bids $b^2 = -3$ to avoid losses. This, however, provides an incentive for player 3 to bid less than -5.99, which in turn provides an incentive for player 2 to overbid the third player and these overbidding incentives only fully vanish when player 3 bids -5.99 again. Because of this circular incentive structure, in equilibrium, player 2 and player 3 will mix strategies. We do not fully characterize the exact mixed strategy equilibrium here, because it is sufficient for our purpose to show that players will not bid in certain intervals.[2]

As outlined before, for player 2, strategies above 0 cannot be part of an equilibrium. Hence player 3 can ensure himself a payoff of at least $Eu_i = 1.495$ by bidding -5.99. Importantly, strategies that are part of a mixed strategy equilibrium must lead to a higher payoff than strategies that are not part of this equilibrium. Hence, bidding $b^3 \in (-5.99, -2)$ cannot be part of a mixed strategy equilibrium because it leads to lower payoffs than bidding -5.99, independent of how player 2 exactly mixes pure strategies below $b^2 = 0$. Bidding $b^3 \in [-8, -5.99)$ could in principle lead to the same payoff (or even a higher payoff) as bidding -5.99 because the commodity's real value is underbid by a larger amount. The same is true for bidding $b^3 \in [-2, -1.50]$ because player 3 might overbid player 4 with these bids. By bidding above -1.5, player 3 might still overbid player 4, but the (maximal) payoff ($Eu_i = 1.49$) resulting from these bids is lower than the payoff of bidding -5.99. Bearing these considerations in mind, player 4 could always avoid to be overbid by player 3 by bidding $b^4 = -7.49$ and ensuring himself a payoff of $Eu^4 = 2.245$. Because player 3, however, does not bid -5.99 as a pure strategy but possibly also mixes strategies over $[-8, -5.99]$ and $[-2, -1.50]$, player 4 potentially mixes strategies over $-8 \leq b^4 \leq -7.49$. Importantly, bidding above -7.49 cannot be part of an equilibrium because then payoffs are lower than $Eu^4 = 2.245$. Especially overbidding player 5 even when this player is bidding $b^5 = -8$ would only lead to an expected payoff of $Eu^4 = 2.0$. For this reason, the influence on strategies of boundary-signals ends at player 5: This player and all players with higher signals than player 5 will play -8 as a pure strategy in equilibrium because their lower-signal opponents do not have an incentive to overbid them. Or in other words,

[1] More precisely, due to the rule we implement concerning equal bids, overbidding in this context means that player 1 only has to bid exactly player 2's absolute bid in order win the auction.

[2] The strategy space in our experiment is finite because participants have to round their bids to the cent-level. But for finite strategy spaces we know that there always exists an equilibrium.

for signals above 46, bidding -8 remains the equilibrium.

Additionally, at the higher boundary of the commodity's value space, no problems occur: A player receiving the signal $x^{high} \in (222, 228]$ knows that the commodity's real value is below his own signal. Hence, he has to underbid his opponent who has a lower signal in order to earn money. But this do not lead to a change in equilibrium because if the opponent bids -8, the player with x^{high} also just bids -8 and has no incentive to deviate.

C.2 Figures

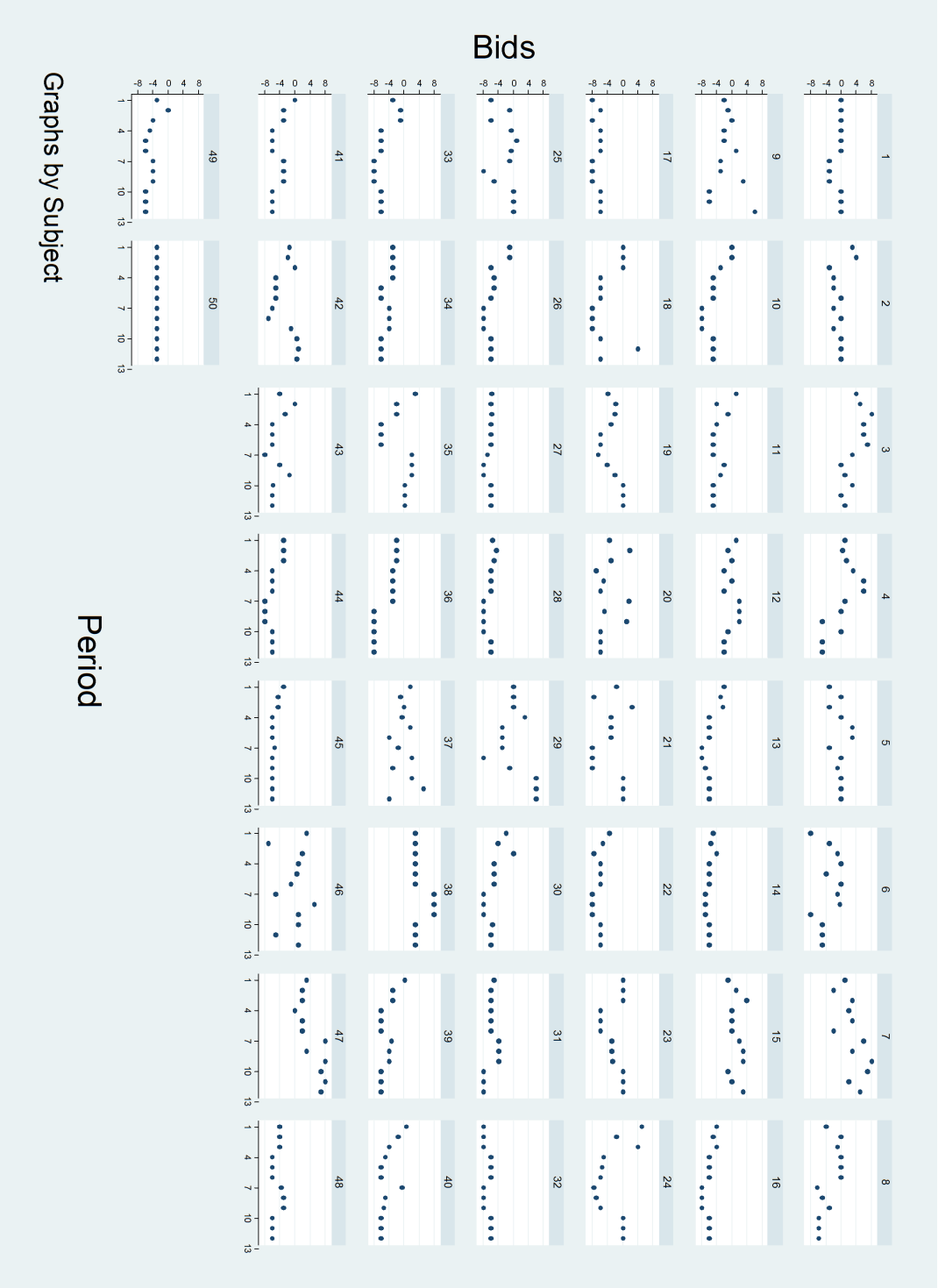


Figure C.1: Each Subjects' Behavior in the AuctionFirst treatment (3 periods per game)

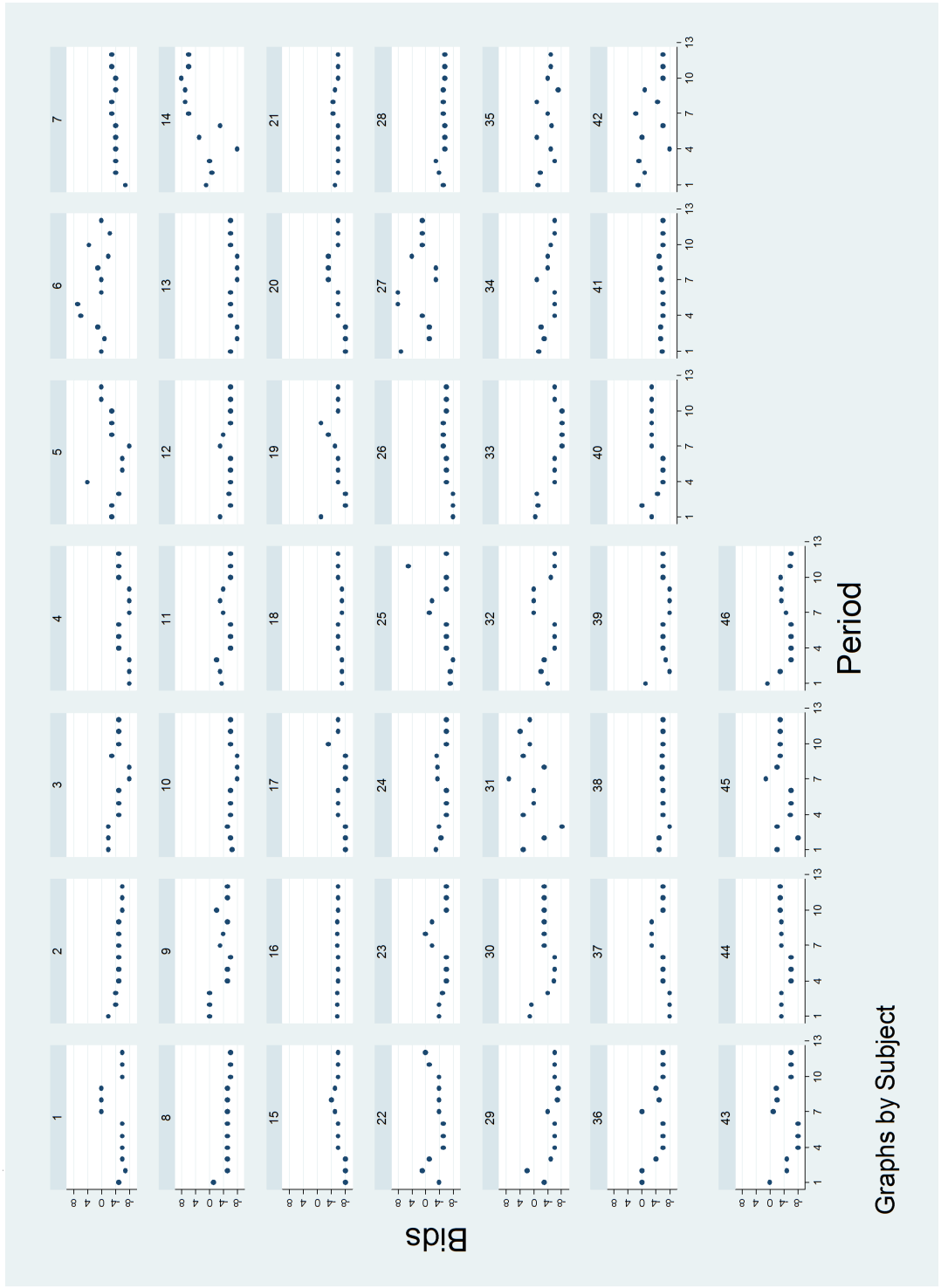


Figure C.2: Each Subjects' Behavior in the TransformedFirst treatment (3 periods per game)

C.3 Instructions: *AuctionFirst* treatment

Welcome to the experiment!

Introduction

I welcome you to today's experiment. The experiment is funded by the University of Mannheim. Please follow the instructions carefully.

For participating, you first of all receive a participation fee of 4€. Additionally, you may earn a considerable amount of money. Your decisions and the decisions of other participants determine this additional amount. You will be instructed in detail how your earnings depend on your decisions and on the decisions of other participants. All that you earn is yours to keep, and will be paid to you in private, in cash, after today's session.

It is important to us that you remain silent and do not look at other people's screens. If you have any questions or need assistance of any kind, please raise your hand, and an experimenter will come to you. If you talk, shout out loud, etc., you will be asked to leave.

The experiment consists of three parts. For all three parts, you will receive separate instructions. You will first make your decisions for all three parts and only afterwards **at the very end** of the experiment get to know which payments resulted from your decisions. The currency used in all three parts of the experiment is called Taler. Naturally, however, you will be paid in Euro at the end of the experiment. **Two Taler will then convert to one Euro.**

If you have any questions at this point, please raise your hand.

Part I

The first part of the experiment consists of 2×3 trading periods (thus trading periods 1-3 and trading periods 4-6). These instructions describe the decision problem as it is present in trading periods 1-3. This decision problem will be slightly modified in the trading periods 4-6. You will be informed about the details of this modification at the end of trading periods 1-3.

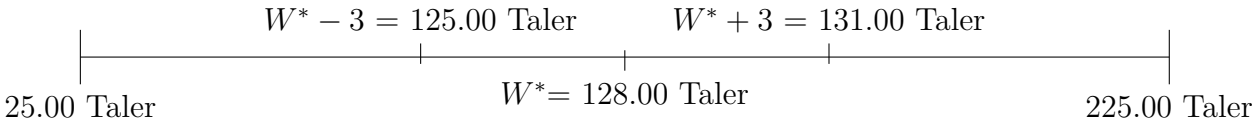
In this part of the experiment, you will act as a buyer of a fictitious commodity. In each trading period, you will have the opportunity to submit a bid for one unit of the commodity. Importantly, not only you will have this opportunity to make a bid for the commodity. In each trading period, you will be matched with another participant of this experiment. This participant will also have the opportunity to make a bid for the commodity. Importantly, you will always bid against another randomly determined participant in each trading period.

Your task is to submit bids for the commodity in competition with the other participant. The precise value of the commodity at the time you make your bids will be unknown to you. Instead, you and the other participant will receive an information signal as to the value of the item which you should find useful in determining your bid. Which kind of information you will receive, will be described below.

The value of the auctioned commodity (W^*) will alwalys be an integer and will be assigned randomly. This value can never be below 25 Taler and never be above 225 Taler. Additionally, the commodity's value W^* is randomly and independently determined from trading period to trading period. As such a high W^* in one period tells you nothing about the likely value in the next period

Private Information Signals: Although you do not know the precise value of the commodity, you and the participant who is matched with you will receive an information signal that will narrow down the range of possible values of the commodity. This information signal is either W^* – 3 or W^* + 3, where both values are equally likely. In addition, it holds that when you receive the information signal W^* – 3, the person who is matched to you will receive the information signal W^* + 3. If in contrast, you receive the information signal W^* + 3, the other person gets the information signal W^* – 3.

For example, suppose that the value of the auctioned item (which is initially unknown to you) is 128.00 Taler. Then you will either receive a) the information signal $W^* - 3 = 125.00$ Taler or b) the information signal $W^* + 3 = 131.00$. In both cases, the other person will receive the opposite information signal, in case of a) the information signal $W^* + 3 = 131.00$ and in case of b) the information signal $W^* - 3 = 125.00$ Taler. The line diagram below shows what's going on in this example.



It also holds that the commodity's value W^* is equal to the signal – 3 or the signal + 3 with equal probability. The computer calculates this for you and notes it.

Your signal values are strictly private information and are not to be revealed to the other person. In addition, you will only be informed about the commodity's value W^* and the other participant's bid at the end of the whole experiment (when also the second and the third part of the experiment are completed).

It is important to note that no participant is allowed to bid less than the signal – 8 and more than the signal + 8 for the commodity. Every bid between these values (including these values) is possible. Bids have at least to be rounded **to one cent**. Moreover, it holds that the participant who submits the higher bid gets the commodity and makes a profit equal to the differences between the value of the commodity and the the amount he or she bids. That is,

- Profit = W^* (128.00 Taler) – higher bid

for the higher bidding person. If this difference is negative, the winning person looses money. If you do not make the higher bid on the item, you will neither make a profit nor a loss. You will earn zero profits. If you and the other participant submit the same bid, the person who received the lower signal will get the commodity and he or she will be paid according to his or her bid.

At the beginning of part I, each individual participant will be given a starting capital credit balance of 8 Taler. Any profit earned by you in the experiment will be added to this sum. Any losses incurred will be subtracted from this sum. At the end of this part of the experiment, all gains and losses will be add up and the net balance of these transactions will be added to your captital credit balance. You are permitted to bid in excess of your capital credit balance. Even in case of a negative captial credit balance, you are still permitted to submit bids. Should your net balance at the end of this part of the experiment be zero (or less), you will not get any payoff from this part of the experiment. But even in case you make losses in this part of the experiment, you will keep your initial show-up fee of 4€.

Summary:

1. Two participants have the opportunity to submit bids for a fictitious commodity. The exact value of the commodity W^* is unknown to you. This value will, however, always be between 25 Taler and 225 Taler. Moreover, you receive a private information signal concerning the commodity's value. This signal is either $W^* - 3$ or $W^* + 3$. The other participant will receive the other signal. No one is allowed to bid less than the signal – 8 or more than the signal + 8.

2. The higher-bidding participant gains the commodity and makes the following profit = commodity's value - higher bid.

3. Profits will be added to your initial capital starting balance. Losses will be subtracted from your initial capital starting balance. You can always submit higher bids than your capital starting balance.

4. This part of the experiment consists of two rounds with overall 6 trading periods. These instructions describe the decision problem as it occurs in the trading periods 1-3. There will be a modification of the decision problem for rounds 4-6, about which you will be informed soon.

If you have read everything, please click the "Ready" button, to start the experiment.

Modifciation of the decision problem

You have now entered all decisions for the trading periods 1-3. Now, trading periods 4-6 will follow for which the decision problem so far will be slightly modified. As up to now the task is to submit bids for a fictitious commodity. Importantly, the other participant who also has the opportunity to submit bids will be replaced by the computer. As the other participant in the trading periods 1-3, the computer will also receive a signal about the commodity's value that is opposite to your own signal. The computer then decides according to the following decision rule: ***The computer always exactly bids his information signal***. Suppose, for example,

that the true value of the commodity is 128.00 Taler. If the computer receives the information signal 125.00 Taler (commodity's value – 3), the computer's bid is equal to 125.00 Taler. If the computer receives the information signal 131.00 Taler (commodity's value + 3), the computer's bid is equal to 131.00 Taler. Otherwise, everything else does not change.

If you have read everything, please click the "Ready" button, to continue with the experiment.

Part II

The second part of the experiment consists of 3 trading periods (trading periods 7-9). In this part of the experiment, you will again act as a buyer of a fictitious commodity. In each trading period, you will have the opportunity to submit a bid for one unit of the commodity. Importantly, not only you will have this opportunity to make a bid for the commodity. In each trading period, you will be matched with another participant of this experiment. This participant will also have the opportunity to make a bid for the commodity. Importantly, you will always bid against another randomly determined participant in each trading period.

Your task is to submit bids for the commodity in competition with the other participant. In general, the value of the auctioned commodity will always be an integer and will be randomly determined. This value can never be below 25 Taler and never be above 225 Taler. At the beginning of each period, you and the other participant will be informed about the commodity's value. Importantly, however, there is a slight uncertainty about the value of the commodity. This value can take two different specifications in every period. The commodity can either be worth W_1^* or W_2^*, where both values always differ by 6 Taler and W_1^* always indicates the lower value. Which of the two values really realizes depends on chance and your bid as well as the other participant's bid and will be explained to you in more detail below. Both your bid and the other participant's bid are not allowed to be lower than $W_1^* - 5$ or higher than $W_2^* + 5$. Every bid between these values (including these values) is possible. Bids have at least to be rounded **to one cent**.

To make the rules of the auction understandable, they will be explained in detail with the help of an example. Suppose that at the beginning of one period, you are informed that the commodity's value is either $W_1^* = 107.00$ Taler or $W_2^* = 113.00$ Taler. You and the other participant are not allowed to bid less than $W_1^* - 5 = 102.00$ or more than $W_2^* + 5 = 118.00$ Taler. Who gets the commodity depends on your bid and the other participant's bid. Three rules apply:

1. **Your bid is 6.00 Taler or more higher than the other participant's bid:**
   In this case, you will get the commodity for sure. With a 50 percent chance each the commodity's value then is either W_1^* (107.00 Taler) or W_2^* (113.00 Taler). Hence, your profit is:

- Profit = W_1^* (107.00 Taler) – Your bid or
- Profit = W_2^* (113.00 Taler) – Your bid

Both scenarios are equally likely and the computer will randomly choose which scenario occurs. If one of the differences is negative and this scenario occurs, you will make a loss. The other participant will be paid according to rule 2.

2. **Your bid is 6.00 Taler or more below the other participant's bid:**
   In this case, you will not get the commodity in any case and your profit is zero. The other participant will be paid according to rule 1.

3. **Your bid is less than 6.00 Taler above or less than 6.00 Taler below the other participant's bid:**
   In this case, either you or the other participant get the commodity with a 50 percent chance and the computer will make this decision. The commodity's value is in any case W_1^* (107.00 Taler). Hence, in case you get the commodity, your profit is:

   - Profit = W_1^* (107.00 Taler) – Your bid

   In this case, the other participant earns zero Taler. If on the contrary, you do not get the commodity, your profit is zero and the other participant's profit is:

   - Profit = W_1^* (107.00 Taler) – His/her bid

   In both cases, it holds for the person who gets the commodity that this person will make a loss if the difference is negative.

At the beginning of part II, each individual participant will be given a starting capital credit balance of 8 Taler. Any profit earned by you in the experiment will be added to this sum. Any losses incurred will be subtracted from this sum. At the end of this part of the experiment, all gains and losses will be add up and the net balance of these transactions will be added to your captital credit balance. You are permitted to bid in excess of your capital credit balance. Even in case of a negative captial credit balance, you are still permitted to submit bids. Should your net balance at the end of this part of the experiment be zero (or less), you will not get any payoff from this part of the experiment. But even in case you make losses in this part of the experiment, you will keep your initial show-up fee of 4€.

You will only be informed about the other participant's bid and which value of commodity actually has realized at the end of the whole experiment (when also the third part of the experiment is completed).

Summary:

1. Two participants have the opportunity to submit bids for a fictitious commodity. The value of commodity will always be between 25 Taler and 225 Taler. Because of uncertainty, the commodity's value can take two specifications W_1^* and W_2^*, where the difference between both values is always 6 Taler. No one is allowed to bid less than W_1^* – 5 and more than W_2^* + 5.

2. If one person bids at least 6.00 Taler more than the other person, this persons gets the commodity for sure and either makes the profit = W_1^* – his/her bid or the profit = W_2^* – his/her bid. If one person bids at least 6.00 Taler less than the the other person, this person does not get the commodity in any case and makes a profit of zero Taler. If the difference of the bids is less than 6.00 Taler, both participants get the commodity with a 50 percent chance and make the following profit = W_1^* – his/her bid in this case.

3. Profits will be added to your initial capital starting balance. Losses will be subtracted from your initial capital starting balance. You can always submit higher bids than your capital starting balance.

4. This part of the experiment consists of 3 trading periods.

If you have read everything, please click the "Ready" button, to continue with the experiment.

Part III

The third part of the experiment consists of 3 trading periods (trading periods 10-12). These 3 trading periods are almost identical to the trading periods 7-9 of part II. In addition, your capital credit balance of the end of part II will be the starting capital credit balance of this part. Hence, the payoff you receive from part II and part III of the experiment will finally depend on the amount of the capital credit balance at the end of this part of the experiment. In part III of the experiment, the following modification of the decision problem of part II is implemented: As up to now the task is to submit bids for a fictitious commodity. Importantly, the other participant who also has the opportunity to submit bids will be replaced by the computer. As the other participant in the trading periods 7-9, the computer is informed about both possible values of the commodity. The computer then decides according to the following decision rule: ***The computer always exactly bids the mean value of both values of the commodity (hence*** $\frac{W_1^*+W_2^*}{2}$ ***or*** $W_1^* + 3 = W_2^* - 3$***)***. Suppose, for example, that the true value of the commodity is either W_1^* = 107.00 Taler or W_2^* = 113.00 Taler. The computer will then bid 110.00 Taler ($\frac{107+113}{2}$= 107.00 + 3.00 = 113.00 – 3.00). Otherwise, everything else does not change.

If you have read everything, please click the "Ready" button, to continue with the experiment.

C.4 Instructions: Frequently Asked Questions

Auction game

1. *When I make my decision about which bid to submit, what kind of specific information do I have? Do I know the true value of the commodity?*
   You do not know the commodity's value W^*. When making your decision, you only know your private information signal. You also do not know whether you received the "high" or the "low" signal. You only receive one number. With a 50 percent chance, you have received the high signal and with a 50 percent chance you have received the low signal. All this also holds correspondingly for the other participant.

2. *On what does it depend whether I get the commodity and how much do I earn should this situation arise?*
   The person who submits the higher bid gets the commodity. The profit then is: W^* – higher bid. If both bids are exactly the same (meaning bids are also the same on the cent-level), the person with the lower signal gets the commodity.

3. *Which values am I allowed to bid?*
   You are allowed to under- and overbid your personal information signal by up to 8.00 Taler. In addition, it is important that you are not only allowed to bid integers. For example, you could also bid 30.45 Taler instead of 30 Taler.

Transformed treatment

1. *When I make my decision about which bid to submit, what kind of specific information do I have? Do I know the true value of the commodity?*
   When making your decision, you know about two possible specifications of the commodity's value: W_1^* and W_2^*. Which of these values actually realizes in the end depends on your decision, the other participant's decision and chance.

2. *On what does it depend whether I get the commodity and how much do I earn should this situation arise?*
   If you at least bid 6.00 Taler more than the other person, you will get the commodity for sure. Your profit will then be W_1^* – your bid or W_2^* – your bid, with a 50 percent chance each. Conversely it holds, that if you bid at least 6.00 Taler less than the other person, you will not get the commodity and your profit will be zero. If the difference of the bids is smaller than 6.00 Taler, either you or the other participant gets the commodity with a 50 percent chance and the computer will make this decision randomly. If the computer chooses you as the winner, your profit will be W_1^* – your bid.

3. *Which values am I allowed to bid?*
   You are allowed to underbid the lower value of the commodity W_1^* by up to 5.00 Taler

and overbid the higher value of the commodity W_2^* by up to 5.00 Taler. In addition, it is important that you are not only allowed to bid integers. For example, you could also bid 30.45 Taler instead of 30 Taler.

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig angefertigt und mich keiner anderen als der in ihr angegebenen Hilfsmittel bedient zu haben. Insbesondere sind sämtliche Zitate aus anderen Quellen als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, den 6.05.2014 *C. Koch*

Curriculum Vitae

2014 –	Postdoctoral Associate NYU Abu Dhabi
2009 – 2014	PhD Studies in Economics Center for Doctoral Studies in Economics (CDSE) University of Mannheim
2004 – 2009	Diploma in Economics & Mathematics (Master equivalent) *Diplom-Wirtschaftsmathematiker* University of Hamburg
2006 – 2007	Visiting Student (ERASMUS) University of Leicester
1995 – 2004	Abitur Gymnasium Schwarzenbek