Expectations, Preferences, and Social Risk Taking

Von der Fakultät für Wirtschaftswissenschaften der
Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors der
Wirtschafts- und Sozialwissenschaften genehmigte Dissertation

vorgelegt von

Johann Philipp Schade

Berichter: Univ.-Prof. Dr. rer. pol. Wolfgang Breuer
Univ.-Prof. Dr. rer. pol. Rüdiger von Nitzsch

Tag der mündlichen Prüfung: 11.07.2017

Table of Contents

1 Introduction

Many important economic decisions incorporate risk or uncertainty which does not arise from nature in the form of a natural disaster or a health risk, but from the action of another human being. To describe risk and uncertainty caused by human beings rather than by nature, Bohnet, Greig, Herrmann, and Zeckhauser (2008) refer to the term "social risk".[1] Social risk is especially ubiquitous in strategic situations, where one's outcome does not only depend on one's own action but on the action of another person. For example, one might think of bank clients whose wealth depends on the investment strategies of their bank managers, entrepreneurs whose profits depend on whether their customers appreciate and purchase their products, or employees whose future salary depends on whether their supervisors will recommend them for promotion.

According to standard economic theory, social risk arises from the existence of multiple equilibria in pure strategies (Aydogan, 2015). In a pure strategy equilibrium, each player chooses one of the available strategies with certainty, and any unilaterally deviating from an equilibrium strategy would lower her or his utility. If there are multiple pure strategy equilibria, the solution of the game is a priori ambiguous: Players are faced with social risk, because it is uncertain which equilibrium strategy the other players will choose. By implication, social risk

[1] Rigorously speaking, the terms "risk" and "uncertainty" have a slightly different meaning. Decisions can be defined as risky if decision makers do not know the actual outcome of their decisions but do know the probability distributions of the possible outcomes. Uncertainty implies that both the actual outcomes of the decisions and the probability distributions of the possible outcomes are a priori unknown. However, the literature on social risk taking does not clearly distinguish between risk and uncertainty (see, e.g., Ben-Ner & Putterman, 2001; Bohnet & Zeckhauser, 2004; Coleman, 1990). In line with the literature, I will also speak of "risk" even where there is "uncertainty", as long as I do not deem a more rigorous distinction to be relevant.

does not occur in games with a unique pure strategy equilibrium. In this case, the equilibrium perfectly determines the moves of all players and thus the outcome of the game. Social risk caused by multiple pure strategy equilibria can be denoted as strategic risk. In response to strategic risk, players may apply so-called mixed strategies, i.e. they select each of the pure equilibrium strategies with a certain probability. A common example for such a randomization is the hand game "rock, paper, scissors".

Another source of social risk is imperfect information: Decision makers who are not fully aware of their counterparts' preferences may have difficulty predicting the behavior of the latter correctly. Standard economic theory does not allow for this type of social risk, since it assumes that "*all* people are *exclusively* motivated by their material self-interest" (Fehr & Schmidt, 2000). This basic assumption of standard economic theory has been widely challenged in the last decades. Countless studies in the field of economics and psychology have pointed out the relevance of social preferences such as altruism or inequality aversion, which basically imply that individuals are not strictly selfish but willing to sacrifice their own payoff to increase the payoff of others (see, e.g., Kagel & Roth, 2009 for an overview). As social preferences may differ significantly between individuals, e.g. depending on gender (e.g. Croson & Gneezy, 2009) or on culture (e.g. Buchan, Johnson, & Croson, 2006), their existence gives rise to social risk in any strategic situation, even where there is a unique pure strategy equilibrium.

Recent experimental studies emphasize the need for a better understanding of social risk taking and the underlying behavioral mechanisms. In fact, individuals treat social risk and natural risk differently (Bohnet & Zeckhauser, 2004). This is perhaps the reason why social risk taking behavior cannot be fully explained by attitudes toward conventional risk (see Trautmann & Vieider, 2012 for an overview). The present thesis studies social risk taking in two

specific situations: a two-player market entry game (Kahneman, 1988; Selten & Güth, 1982) and a binary-choice trust game (Bohnet & Zeckhauser, 2004). The main characteristics of both games are summarized in Table 1. As will be outlined in more detail, there are some general differences. However, both games constitute a similar decision problem, because players have to choose between a safe and a risky option.

	Number of Players	Strategies per Player	Pure Strategy Equilibria	Is Game Played Sequentially?
Two-Player Market Entry Game	2	2	2	no
Binary-Choice Trust Game	2	2	1	yes

Table 1: Characteristics of the games studied

The two-player market entry game can be framed as an entrepreneurial decision problem, where two firms have to decide simultaneously whether to enter a market or not. Since the market capacity is limited, e.g. due to constraints on the demand side, a market entry is successful if and only if the entrant firm is the only one in the market. The unique market entrant benefits from operating the market, whereas the non-entrant neither earns nor loses money. Simultaneous market entries cause financial harm to both firms. Thus, staying out of the market is the safe option, yielding neither a gain nor a loss, whereas a market entry bears the risk of being competed against and losing money. The payoff matrix of the game, including the payoffs used throughout this thesis, is given in Table 2. A unique market entrant earns 3.5 experimental currency units (ECUs). If both firms enter the market simultaneously, both lose 3.5 ECUs. Staying out of the market yields zero ECUs, irrespective of the other firm's choice.

		Firm B	
		Stay Out	Enter
Firm A	Stay Out	(0, 0)	(0, 3.5)
	Enter	(3.5, 0)	(-3.5, -3.5)

Table 2: Two-player market entry game

Formally, the two-player market entry game has two pure strategy equilibria, where each player chooses to enter the market if the other player chooses to stay out, and vice versa. Due to strategic risk caused by multiple pure strategy equilibria, there is also a mixed strategy equilibrium, where each option is chosen with a probability of 1/2. Thus, social risk arises from the existence of both multiple pure strategy equilibria and social preferences.

Whereas the two-player market entry game is a simultaneous game, the binary-choice trust game is played sequentially. Its payoff matrix, showing the payoffs used throughout this thesis, is given in Table 3. The first mover (*trustor*) has to choose between the "trust" option and the "no trust" option. If the "no trust" option is chosen, both players receive an equal amount with certainty (10 ECUs). In this case, the role of the second player (*trustee*) is limited to the mere reception of money. However, if the "trust" option is chosen, the players' payoffs are determined by the trustee's choice. Speaking in terms of trust, the trustee can either betray or reward the trustor's trust. Reward generates 15 ECUs for both players, whereas betrayal yields 22 ECUs for the trustee and 8 ECUs for the trustor. Thus, the trustor is faced with the decision problem of whether to choose the safe "no trust" option, yielding 10 ECUs with certainty, or the risky "trust" option, where she or he earns either a higher (15 ECUs) or a lower amount (8 ECUs) compared to the safe "no trust" option. Irrespective of the trustee's choice,

the "trust" option yields a higher total payoff (30 ECUs) than the "no trust" option does (20 ECUs). From the "social" perspective, it is therefore beneficial to trust.

		Trustee	
		Reward	Betray
Trustor	Trust	(15, 15)	(8, 22)
	No Trust	(10, 10)	(10, 10)

Table 3: Binary-choice trust game

Under standard assumptions, there is a unique equilibrium in pure strategies, which can be derived by means of backward induction. Rational and selfish trustees should not choose to reward trust. By implication, trustors who perfectly anticipate their trust being betrayed should not even select the "trust" option. Since a unique pure strategy equilibrium exists, there is no strategic risk from the perspective of standard economic theory. Hence, social risk can only arise if social preferences are assumed.

Throughout this thesis, I address three general research questions (RQs). First, I study how decision makers form expectations about their counterpart's behavior in the two-player market entry game as well as in the binary-choice trust game (RQ #1). Depending on the context, expectations refer either to the predicted behavior of the other ("How *will* my counterpart behave?") or to perceived obligations of the other ("How *should* my counterpart behave?"). Second, I investigate how expectations and preferences are mapped into decisions (RQ #2). Third, I examine the interdependencies between expectations and preferences and

their joint influence on decisions (RQ #3). Figure 1 embeds the logical interdependencies between the three RQs in a general framework, which is refined step by step in this thesis. The numbers depicted in Figure 1 refer to the respective RQs.

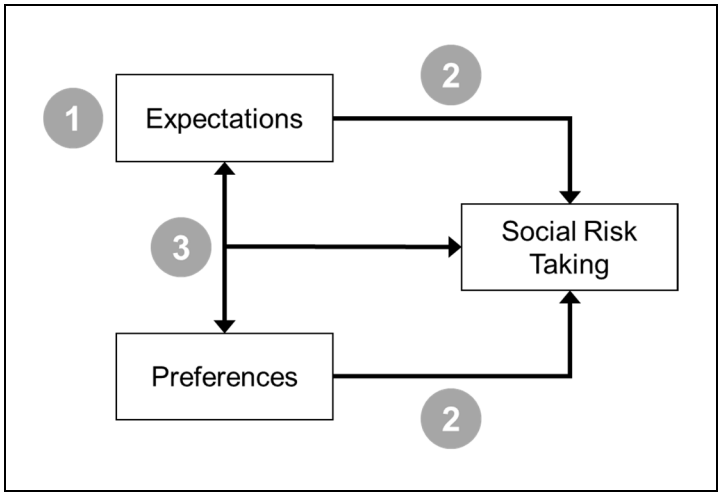

Figure 1: Interdependencies between research questions

The present thesis comprises four essays, which are formally independent of each other. The numeration of the essays takes into account the respective game on which the respective essay is based: Whereas Essay #1 examines the two-player market entry game, Essays #2.1, #2.2, and #2.3 build on the binary-choice trust game. Although each essay focuses on a specific aspect of social risk taking, using two different games, they largely overlap with respect to the general framework developed above. In the following, I briefly review the four essays and indicate to which of the RQs they refer:

– Essay #1, which is entitled "The Twofold Beauty Premium: Preference-Based and Beliefs-Based Behavior in an Anti-Coordination Game", reveals three results.[2] First, individuals taking part in a two-player market entry game form beliefs about their opponents' behavior based on stereotypes (RQ #1), which relate people's physical appearance to their willingness to take social risk. Second, people seem to refrain from market entry because they have a preference for good-looking counterparts. Both beliefs-based and preference-based behavior is moderated by people's gender (RQ #3).

– Analogously to Essay #1, the implications of Essay #2.1, entitled "The Higher Your Expectations, the Lower Your Trust: Avoiding the Experience of Unfulfilled Expectations", are threefold.[3] Studying a binary-choice trust game, my co-authors and I find evidence for a consensus effect, which generally implies that people tend to extrapolate from their preferences to the believed preferences of others (Mullen et al., 1985; Ross, Greene, & House, 1977). In line with this concept, trustors form beliefs about trustees' behavior based on their own social preferences (RQ #1). Moreover, trustors who are optimistic with respect to their counterparts' trustworthiness lower their willingness to trust in order to avoid their optimistic expectations being disconfirmed (RQ #2).

– In Essay #2.2, which is entitled "Breaking the Rules: Anticipation of Norm Violation in a Binary-Choice Trust Game", my two co-authors and I investigate the role of moral

[2] Essay #1 has been presented at the 2015 "Academic Summer" conference held by the Management Academy of the German Savings Banks Association (*DSGV*) in Bonn as well as at the 2015 annual meeting of the German Association for Experimental Economic Research (*GfeW*) in Hamburg.

[3] Essay #2.1 has been presented at the 2016 annual meeting of the German Association of Experimental Economic Research (*GfeW*) in Gießen.

expectations in a binary-choice game (RQ #2).[4] More precisely, we consider the obligations which trustors assign to the trustee role. Our study reveals that trustors who consider trustees to be obligated to act trustworthily are, ceteris paribus, less willing to trust. The underlying reason may be that trustors attempt to hinder trustees from violating their moral obligation to reward trust.

– In Essay #2.3, entitled "Once Bitten, Twice Shy: Trust, Trustworthiness, and the Hot-Cold Empathy Gap", my co-author and I study a repeated binary-choice trust game in which players take the trustor role and then take the trustee role, and vice versa. We find that people's beliefs about others' trustworthiness (i.e. whether others will reward or betray trust) moderate the intrapersonal correlation between trust and trustworthiness (RQs #2, #3). If subjects hold optimistic beliefs, their willingness to trust is positively related to their willingness to reward. We find a contrary relationship for pessimistic subjects.

The remainder of the present thesis is structured as follows. Section 2 is dedicated to placing the two-player market entry game and the binary-choice trust game in a broader game-theoretical context. The binary-choice trust game is part of the trust game family, whereas the two-player market entry game constitutes a so-called anti-coordination game. Both classes of games are briefly introduced in Section 2. In Section 3, the role of social preferences is discussed with respect to the two-player market entry game and the binary-choice trust game. Section 4 provides an integrated view of the essays, which are included in Section 5.

[4] Essay #2.2 has been published in *Economics Letters* (Breuer, Helduser, & Schade, 2016).

The essays are all similarly structured. First, the research question and its relevance are outlined. Second, hypotheses are derived from the related literature. The third step provides insights into the methods used. As all essays are experimental studies, this section especially features the experimental design and the procedures of the respective experiments. Next, the results obtained from the experiments are presented and discussed with respect to the hypotheses. The results section is followed by a conclusion, covering theoretical and practical implications as well as open research questions.

2 Social Risk in Two Classes of Games

2.1 Anti-Coordination Games

The two-player market entry game studied in Essay #1 belongs to the class of anti-coordination games. Formally, anti-coordination problems refer to situations where it is mutually beneficial for players to choose opposite strategies. The difficulty to coordinate one's actions arises from the existence of multiple pure strategy equilibria. As a rational player cannot anticipate for which of the equilibria the other player will opt, it is difficult to choose the optimal response to the other player's choice. The two-player market entry game has two equilibria in pure strategies: Firm A enters the market if Firm B stays out, and vice versa. In neither of the two equilibria does any player have an incentive to choose a deviating strategy. The existence of two pure strategy equilibria results in an additional equilibrium in mixed strategies, implying that players will choose either of the two strategies at hand with a probability of $1/2$ (given the specific payoffs shown in Table 2).

In the context of multiple pure strategy equilibria games, the role of risk attitudes has been studied with respect to coordination games only. In contrast to anti-coordination games, the pure strategy equilibria of coordination games imply that players choose the same or corresponding strategies. Heinemann, Nagel, and Ockenfels (2009) and Büyükboyacı (2014) investigate risk taking in coordination games and do not find a significant correlation between risk attitudes and coordination behavior. The stag-hunt game studied by Büyükboyacı (2014) is structurally similar to the two-player market entry game, because it also involves a safe and a risky option. An arbitrarily chosen payoff matrix of the stag-hunt game is given in Table 4. The same as in the two-player market entry game, each player has to choose between a safe option ("hunt hare") and a risky one ("hunt stag"). If a player decides to hunt hare, her or his payoff is 3 ECUs. Hunting stag yields either 4 ECUs or 0 ECUs, depending on the other player's choice. Thus, the game has two pure strategy equilibria, where both players simultaneously choose to hunt either stag or hare. In addition, there is a mixed strategy equilibrium, where hunting hare is chosen with a probability of 3/4 (given the specific payoffs shown in Table 4).

		Player 2	
		Hunt Stag	Hunt Hare
Player 1	Hunt Stag	(4, 4)	(0, 3)
	Hunt Hare	(3, 0)	(3, 3)

Table 4: Stag-hunt game

Even if risk preferences do not directly affect players' behavior, players may expect their counterparts to be concerned with risk considerations. Such expectations may cause a second-order effect of risk attitudes. In the experiment by Büyükboyacı (2014), players are provided with information about their counterparts' risk attitudes. Interestingly, players coordinate to the safe strategy ("hunt hare") if they are informed that the other player is risk averse. This finding underlines that people themselves perceive risk attitudes to be a relevant factor in the stag-hunt game. In this vein, the study by Büyükboyacı (2014) also gives rise to the question of whether stereotypes related to other people's risk attitudes affect one's own behavior. This issue is studied in more depth in Essay #1.

2.2 Trust Games

The binary-choice trust game introduced in Section 1 is closely related to the continuous trust game of Berg, Dickhaut, and McCabe (1995), which has received considerable attention in experimental and behavioral economics during the last decades. In a meta-analysis, Johnson and Mislin (2011) review 162 repetitions of this game, involving more than 23,000 participants. For economists, trust games are a reliable measure of trust at the individual level, because they allow the elicitation of individual preferences in a monetarily incentivized way. As such, trust games are expected to be less prone to response biases than, for instance, survey measures are. Another advantage of trust games might be that standard economic theory has a clear prediction of how rational people should behave in trust games. Assuming rationality and self-centered preferences, individuals should not exhibit trust in equilibrium. However, countless repetitions of the trust game have shown that individuals deviate from this theoretical benchmark:

They send and return positive amounts, even when large amounts are at stake (Johansson-Stenman, Mahmud, & Martinsson, 2005).

The importance of trust is widely undisputed in economics as well as in other social sciences. Arrow (1972, p. 357) stated that "virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence." Since then, numerous studies have provided support for Arrow's seminal conjecture: Trust has been shown to reduce transaction costs, to increase the economic performance of firms and governments, and to promote economic growth (e.g. Bromiley & Cummings, 1995; Knack & Keefer, 1997; La Porta, Lopez-de-Silanes, Shleifer, & Vishny, 1997; Zak & Knack, 2001). Given the broad scope of trust, it is not surprising that a multitude of definitions of trust have emerged over the last decades (see, e.g., McKnight & Chervany, 2001 for an overview). Trust games are intended to measure trust according to the definition of trust given by Coleman (1990), who posited three conditions for the existence of trust. First, the decision of whether to trust has to be a risky one. Second, rewarding trust has to be costly for the trusted party. Third, trusting has to increase the total payoff, i.e. the sum of the trusting party's and the trusted party's payoff. The continuous trust game by Berg et al. (1995) as well as the binarized version used in the present thesis (Bohnet & Zeckhauser, 2004) meet all three requirements.

Numerous scholars have posited a close relationship between trust and risk (see Mayer, Davis, & Schoorman, 1995 and references herein). In line with this view, Coleman (1990) argues that "situations involving trust constitute a subclass of those involving risk". Thus, it is not surprising that the role of risk attitudes in the context of trust has attracted much interest.

The related literature can be subdivided into two broad categories. The first category focuses on correlations between risk attitudes and trusting behavior at the individual level. Studies falling into this category elicit trust and risk attitudes by using different concepts to measure each variable, e.g. the continuous trust game as a measure of trust and the multiple price list approach (Holt & Laury, 2002) for measuring risk aversion. The evidence reported in this literature is mixed. Eckel and Wilson (2004) run a binary-choice trust game and implement behavioral measures as well as survey measures to elicit subjects' risk attitudes. None of these instruments reveals a significant correlation between individual risk attitudes and the decision to trust. However, other scholars argue that the decision whether to trust is ambiguous rather than risky (Corcos, Pannequin, & Bourgeois-Gironde, 2012; Houser, Schunk, & Winter, 2010), because the probabilities of trust being rewarded or trust being betrayed are a priori unknown to decision makers (see footnote on p. 1). Therefore, Corcos et al. (2012) elicit ambiguity preferences by means of a multiple price list method, where subjects have to choose between risky and ambiguous lotteries. They find a significant negative correlation between ambiguity aversion and trust. A third contribution stands out with respect to the subject pool. Whereas the participants of the Eckel and Wilson (2004) and the Corcos et al. (2012) experiments are college students, Schechter (2007) conducts a trust game among rural villagers from Paraguay. Also using a lottery choice task to elicit risk attitudes, she finds a significant negative correlation between subjects' risk aversion and their willingness to trust. This result is remarkable, as the amounts at stake are substantially higher compared to Eckel and Wilson (2004) and Corcos et al. (2012).

The second branch of the literature has been largely inspired by the seminal contribution of Bohnet and Zeckhauser (2004), who propose a novel between-subjects design to disentangle

the influence of risk preferences on trust from additional social motives. The approach is based on comparing decisions made in two treatments: a binary-choice trust game (which has been briefly introduced in Section 1) and a risky dictator game. In both games, subjects acting as trustors have to choose between a safe and a risky option, and their decisions determine their own outcome and the outcome of a second player. In the baseline treatment, denoted as the "trust game", the outcome of the risky option is determined by the second player. The extensive form is depicted in Figure 2. In the control treatment, denoted as the "risky dictator game", the human agent is replaced by a lottery. It is ensured that both treatments yield the same probability distributions over the safe and the risky option.

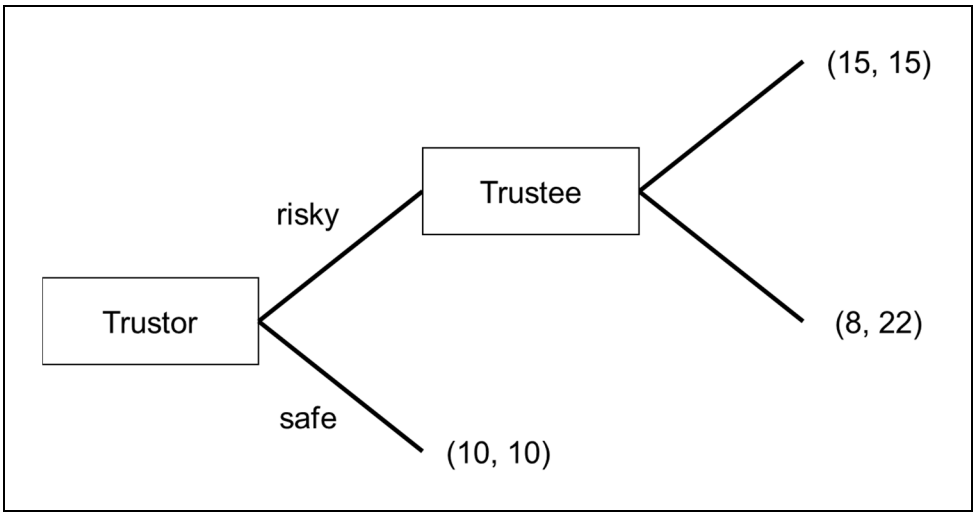

Figure 2: Binary-choice trust game

Another feature of the Bohnet and Zeckhauser (2004) design is that subjects' choices are disentangled from subjects' beliefs about the probability distribution of the risky option (Bohnet et al., 2008; Bohnet, Herrmann, & Zeckhauser, 2010). Rather than simply choosing the safe or the risky option, subjects are asked to state the probability at which they would prefer the risky to the safe option, i.e. their so-called *minimum acceptable probability* (*MAP*).

An MAP below or equal to the actual reward probability rp^a indicates that the risky option is chosen, whereas $MAP > rp^a$ indicates the contrary. The actual reward probability rp^a is derived from second players' actual choices in the "trust game" treatment. The procedures are outlined in more detail in Essays #2.1, #2.2, and #2.3, which all build on the binary-choice trust game by Bohnet and Zeckhauser (2004).

The experiments conducted by Bohnet and Zeckhauser (2004) confirm the hypothesis that individuals treat social risk in a different manner than conventional risk. Specifically, individuals state higher *MAPs* in the "trust game" than in the "risky dictator game", i.e. they demand a social risk premium. From this finding, the authors conclude that individuals are betrayal averse: They seem to attribute an additional disutility to the fact that an unfavorable outcome is caused by a human being and not by a randomizing device. The betrayal aversion phenomenon has been replicated in several other studies (Aimone & Houser, 2012; Fetchenhauer & Dunning, 2009, 2012), and has been shown to be persistent across cultures (Bohnet et al., 2010; Bohnet & Zeckhauser, 2004).

3 Social Preferences: A Twofold Source of Social Risk Taking

Understanding social risk taking is closely tied to the question of how decision makers take into account the outcomes of their counterparts in social interactions. Countless experiments have demonstrated that decision makers do not only pursue their own material interests, even if the situation is one-shot and complete anonymity is ensured. In contrast to the assumptions of standard economic theory, individuals are sensitive to the way their decisions affect others,

and they are frequently willing to sacrifice their own payoff in order to increase others' payoffs. A striking example is the behavior which can be typically observed in the dictator game (Kahneman, Knetsch, & Thaler, 1986). In this non-strategic two-player game, one player is endowed with an amount of money and then has to decide how to split this amount between herself or himself and the second player. The recipient is fully passive and cannot affect the outcome of the game. When running the dictator game in the laboratory, researchers typically observe that first movers send substantial amounts (see Engel, 2011 for a survey), which is in clear conflict with the self-interest hypothesis put forward by standard economic theory. Similar deviations can be observed in strategic interactions, such as the trust game.

In the light of the vast empirical evidence of non-selfish behavior, individuals are said to have social preferences (or other-regarding preferences). Empirically, social preferences can be studied by means of experimental games. Levitt and List (2007) provide a summary of how the outcome of the most popular games can be interpreted in terms of social preferences. Formal models designed to explain behavior in these games typically capture social preferences by including a social component in the utility function (e.g. Bolton & Ockenfels, 2000; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006). The theories of social preferences fall into two broad categories (Charness & Rabin, 2000). One category focuses on the distribution of payoffs, i.e. individuals are assumed to have preferences about how payoffs are distributed among themselves and others. Such outcome-based preferences may be further specified, e.g. as altruism, as inequality aversion, or as maximin preferences. Altruism implies that a person's utility increases with the well-being of others. It is also referred to as a form of unconditional kindness (Fehr & Fischbacher, 2002). Inequality aversion denotes one's willingness to sacrifice one's own payoff in order to reach equality in outcomes between oneself and another person.

Maximin preferences refer to one's desire to maximize the minimal outcome in a group (Engelmann & Strobel, 2004). A second category of preferences takes into account the importance of intentions (Rabin, 1993), which can be defined as "beliefs about why an agent has chosen a certain action" (Fehr & Schmidt, 2006). A type of preference belonging to this category is that of reciprocity. According to the intentions-based approach, reciprocity implies that people respond to a prior kind or unkind action in the same kind or unkind manner. However, reciprocity is a hybrid concept, since it can also be the result of outcome-based preferences. For example, one might think of the trustee in the binary-choice trust game (see Section 1 for further details), who has to decide whether to reciprocate the trustor's trust ("reward" option) or not ("betrayal" option). The trustee may decide to reciprocate for two reasons. One the one hand, she or he may simply want to increase the trustor's payoff or to achieve payoff equality (outcome-based preference component). On the other hand, she or he may want to "reward" the trustor for her or his willingness to trust (intentions-based preference component).

The assumption of individuals holding social preferences is not novel. Several influential economists, such as Smith (1759/1982), Becker (1974), Arrow (1981), Samuelson (1993), and Sen (1995), have stressed that maximizing individual welfare is not the only driver of economic behavior (Fehr & Schmidt, 2000). Nevertheless, it is still standard in economics to assume self-interested decision makers. Fehr and Schmidt (2000) suggest two reasons for the persistency of the self-interest hypothesis. A first reason refers to the predictive power of the latter, which has been demonstrated in many contexts. A second reason is methodological: Assuming non-standard preferences bears the risk of becoming arbitrary, because any finding can be theoretically founded if the underlying assumptions on preferences are brought in line with it. Therefore, assuming standard preferences has become a convention among economists.

According to the literature, social preferences may be related to social risk taking in at least two ways, which are both illustrated in Figure 3. The arrow labeled "1" refers to a direct effect of social preferences on social risk taking, whereas the two arrows labeled "2" represent an indirect effect. As the essays included in this thesis demonstrate, this general framework is not complete and can be refined in several aspects. The way how social preferences directly affect social risk taking (see arrow labeled "1") depends on the type of preferences to be considered. In the two-player market entry game, positive outcome-based preferences, such as altruism, should induce a lower willingness to enter the market for two reasons. First, a market entry would cause harm to the opponent if the latter also decided to enter. Second, staying out of the market enables the other player to make a profit. In the binary-choice trust game, a socially orientated trustor should be more willing to put trust in the other person compared to a selfish trustor, because trusting necessarily increases the trustee's payoff, irrespective of whether the latter chooses to reward or to betray. Thus, in both games, positive outcome-based preferences are oppositely related to social risk taking: Whereas the former raise the willingness to take social risk in the binary-choice trust game, the opposite is true in the two-player market entry game. Intentions-based preferences, by contrast, may induce different direct effects of social risk taking than outcome-based preferences do, as will be outlined in more detail in Section 4.2.

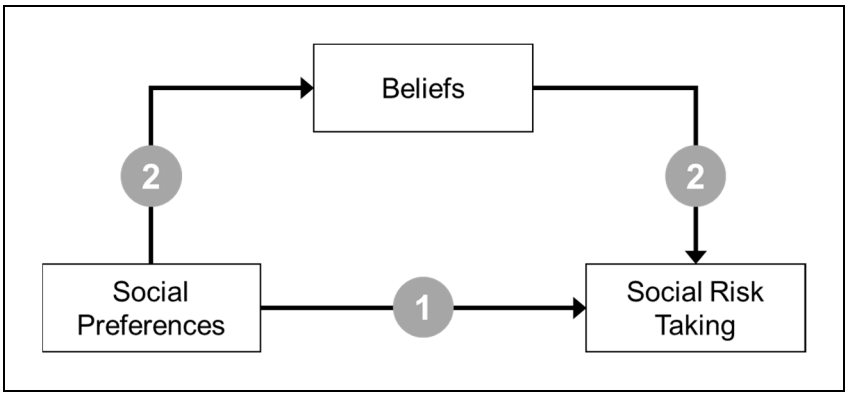

Figure 3: Connections between social preferences and social risk taking

The indirect connection between social preferences and social risk taking (see arrows labeled "2") involves people's beliefs about their counterparts' actions as a third variable. Formally speaking, social preferences may affect social risk taking through the mediation of beliefs (Altmann, Dohmen, & Wibral, 2008; Blanco, Engelmann, Koch, & Normann, 2014). It is in line with intuition that in strategic situations individuals rely on their beliefs concerning the action of others (see arrow labeled "2" on the right). Costa-Gomes, Huck, and Weizsäcker (2014) explicitly show that there is a causal relationship between beliefs and actions in the trust game. However, it is less obvious as to why preferences should affect beliefs (see arrow labeled "2" on the left). Standard economic theory typically assumes independency between preferences and beliefs, but this assumption conflicts with the empirical evidence that individuals rely on their own preferences when predicting the preferences of others (Mullen et al., 1985; Ross et al., 1977). This so-called consensus effect evokes a correlation between beliefs and preferences, and establishes thus an indirect link between social preferences and social risk taking (see arrows labeled "2"), which is entangled with the direct effect of social preferences on social risk taking (see arrow labeled "1").

4 Integrated View of the Essays

4.1 Overview

In Essay #1, I study the interplay between physical attractiveness and gender in a two-player market entry game (Kahneman, 1988; Selten & Güth, 1982). Especially, I examine whether subjects' behavior is affected by the physical appearance of the other participants. According to the literature, the effect of appearance on behavior may be carried out through two channels. First, people may be prone to stereotypes related to physical appearance and risk taking. Based on evolutionary considerations, I hypothesize that good-looking people are considered to be more willing to take social risk, and that people taking part in a two-player market entry game rely on this stereotype when predicting their opponent's behavior. Second, people may have a preference for good-looking counterparts, i.e. a tendency to refrain from market entry in order to not harm the other player or in order to enable the other player to make a profit.

Given the scope of Essay #1, one of the key challenges is to elicit the physical attractiveness of the study participants. I make use of an application designed for mobile electronic devices, which allows the rating of individuals with regard to their facial attractiveness based on symmetry measures. As a second measure, participants are asked to privately judge the other participants' attractiveness.

In line with the literature, I find that co-players' appearance affects individual decision making through two distinct transmission channels: a stereotype-based and a preference-based channel. Both channels are interacted with gender. Females expect other females who are rated as physically attractive to be more willing to enter the market. As subjects behave rationally with regard to their beliefs, good-looking females face fewer market entries in female dyads.

Similarly to females, males enter the market with a lower likelihood when facing a good-looking female co-player. However, males' motivation is preference-based, because their predictions of females' behavior are not affected by females' appearance. Taken together, physically attractive females face a lower risk of being competed against. This applies to both same-sex and mixed-sex dyads, but for different reasons.

In Essay #2.1, my co-authors and I study the role of trustworthiness beliefs in a binary-choice trust game (Bohnet & Zeckhauser, 2004). The term "trustworthiness beliefs" refers to trustors' predictions of trustees' behavior. We hypothesize that trustors decrease their willingness to trust in order to avoid negative emotions associated with unfulfilled expectations, including feelings of disappointment or regret. To test our hypothesis, which largely builds on theoretical and empirical evidence from social psychology, we conduct a laboratory experiment including three parts. First, participants take part in a binary-choice trust game, either in the trustor or in the trustee role. Second, participants have to guess the proportion of trustees who have actually chosen to reward trust. In our analysis, guesses are used as a proxy for trustors' trustworthiness beliefs. Third, participants' social value orientation (Griesinger & Livingston, 1973; Liebrand, 1984) is elicited by means of the *SVO Slider Measure* (Murphy, Ackermann, & Handgraaf, 2011). Generally speaking, the *SVO Slider Measure* serves as a measure of outcome-based preferences (Murphy & Ackermann, 2014). It provides a continuous measure of social value orientation, and allows the classification of individuals into the following four categories (see Essay #2.1 for further details): altruists, competitors, proselfs, and prosocials.

In line with our initial hypothesis, we find that optimism about others' trustworthiness and trust are negatively correlated. This effect is more profound if we control for trustors'

social value orientation in the respective regression model. Taken together, our findings establish an indirect relationship between social value orientation and trust: Social value orientation increases trustors' optimism, which in turn lower one's willingness to trust. The negative correlation between optimism and trust is more pronounced for trustors who expect the trust case to occur. This finding is fully in line with our reasoning, because we expect the respective trustors to be more concerned with the anticipation of negative emotions than others are.

In Essay #2.2, my co-authors and I study the role of perceived obligations in a binary-choice trust game (Bohnet & Zeckhauser, 2004). Our research is largely inspired by Buchan, Croson, and Solnick (2008), who have pointed out the relevance of perceived obligations in the context of trust. However, the scope of Buchan et al. (2008) is limited to the obligations which subjects feel with respect to their own decisions. Our focus, by contrast, lies on the obligations which trustors assign to their counterparts having the trustee role. In addition, we investigate whether trustors behave in line with their obligation statements when taking the trustee role in a repeated binary-choice trust game with reversed roles.

Our main result is straightforward: The higher the obligation which a trustor assigns to the trustee role, the less she or he is willing to trust. We argue that norm-orientated trustors anticipate that they will experience negative emotions should trustees violate their obligation to reciprocate. In order to avoid norm violation, trustors lower their willingness to trust. As a second result, our experiment reveals that trustors behave consistently with their obligation statements when taking the trustee role: Subjects who consider trustees to be obligated to reward trust, reward trust to a greater extent than others do.

In Essay #2.3, my co-author and I study a repeated binary-choice trust game (Bohnet & Zeckhauser, 2004) in which participants take the trustor role and then take the trustee role,

and vice versa. Our objective is to shed light on the relationship between trust and trustworthiness (i.e. whether to reward or to betray trust) within individuals. In line with previous studies, we find that subjects' willingness to trust is positively correlated with their trustworthiness. However, the relationship between trust and trustworthiness is moderated by subjects' trustworthiness beliefs. If subjects hold optimistic trustworthiness beliefs (i.e. they expect trustees to reward trust), their willingness to trust is positively correlated with their own willingness to reward. The contrary relationship applies to pessimistic subjects. We argue that the existence of social preferences, measured in terms of trustworthiness, generates positive and negative emotions with respect to the experience of being rewarded and being betrayed. This implies that a person holding social preferences experiences a higher amount of emotions when being rewarded or being betrayed than a purely selfish person does in either situation. If "social" people place a priori more emphasis on the positive emotions of being rewarded than on the negative emotions of being betrayed, the existence of social preferences induces a higher willingness to trust, and thus a positive correlation between trust and trustworthiness. By contrast, the correlation is negative if people's attention lies on the negative emotions of being betrayed. Based on evidence from psychology, we suggest further that trustworthiness beliefs may determine whether people focus on the emotions of being rewarded or on the emotions of being betrayed. Specifically, optimists concentrate more on the positive emotions of being rewarded than pessimists do.

4.2 Connecting Links

In the present section, I elaborate connecting links between the essays presented above. In fact, each essay reveals specific behavioral mechanisms which can be mapped to the general framework of social risk taking developed in Sections 1 and 3. As outlined in Essay #1, the appearance of one's opponent may affect an individual's market entry decision at two different levels. The influence may be carried out either through a beliefs-based or a preference-based channel. As a second result, Essay #1 reveals that the relevance of the respective channel for actual decision making depends on how female and male participants are matched. Thus, gender serves as a moderator for the relationship between appearance, preferences, and expectations. The two main results obtained from Essay #1 are summarized in Figure 4, which extends the general framework depicted in Figure 1 (see p. 6) and Figure 3 (see p. 19) to the context of the two-player market entry game (except for the direct relationship between beliefs and preferences, which does not fall in the scope of Essay #1). The arrows labeled "1" and "2" denote the beliefs-based and the preference-based channel, respectively. The arrows labeled "3" refer to the moderator effect of gender.

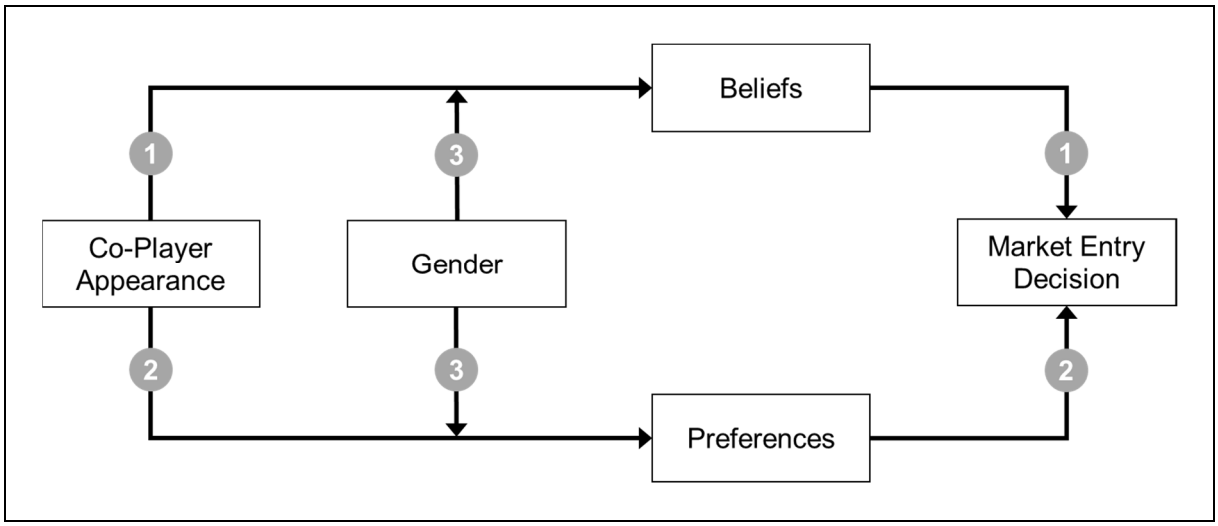

Figure 4: Connections between co-player appearance and market entry decision

Based on a binary-choice trust game, Essays #2.1, #2.2, and #2.3 explore the relationship between social preferences, trustworthiness expectations, and trust. The results from all three essays are summarized in Figure 5, which extends the general framework depicted in Figure 1 (see p. 6) and Figure 3 (see p. 19) to the specific context of trust. Essay #2.1 reveals a negative relationship between trustworthiness beliefs and trust (see arrow labeled "2"), due to the avoidance of anticipated disappointment or regret. As a second result, we demonstrate that this effect is part of an indirect effect, which links social preferences, measured in terms of social value orientation, to trust through the mediation of trustworthiness beliefs (see arrows labeled "1" and "2"). Besides the negative effect of trustworthiness beliefs on trust, the indirect effect builds on a consensus effect inducing a positive relationship between social value orientation and trustworthiness beliefs (see arrow labeled "1"): Socially orientated trustors expect trustees to behave socially, i.e. to be trustworthy, whereas the contrary is true for selfish trustors. The triangular relationship between social value orientation, trustworthiness beliefs,

and trust is completed by a positive direct effect of social value orientation on trust (see arrow labeled "3").

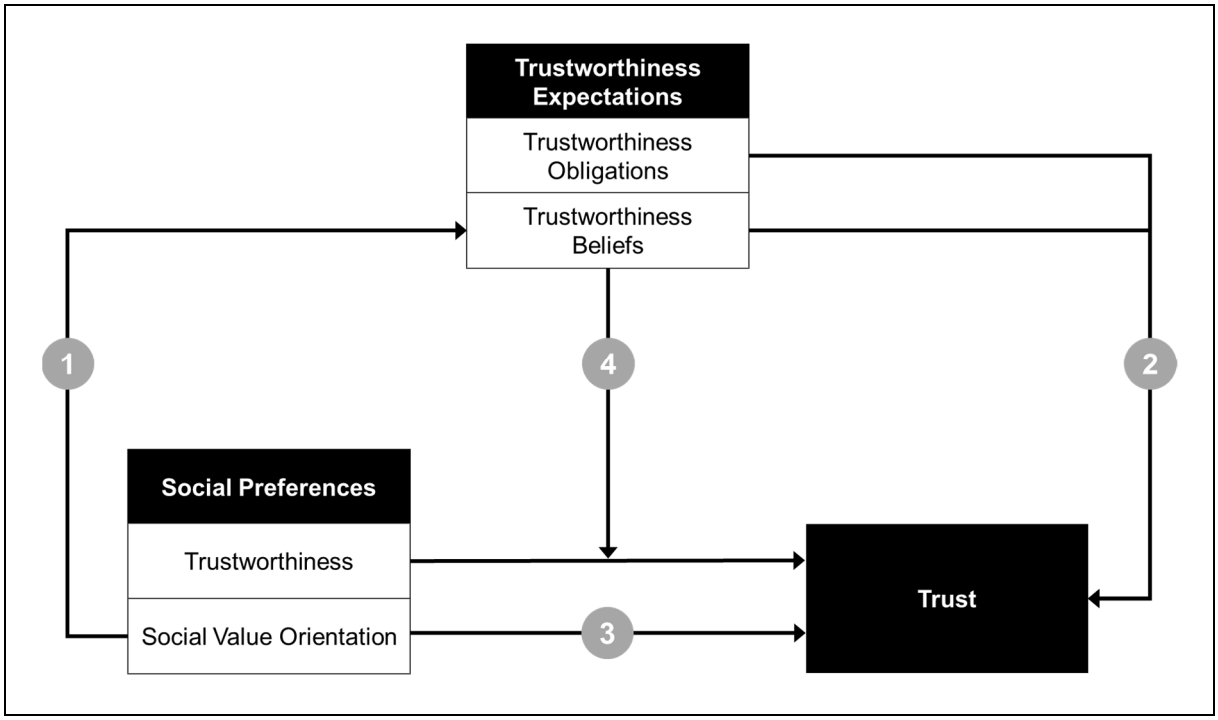

Figure 5: Connections between social preferences, trustworthiness expectations, and trust

Whereas Essay #2.1 investigates the role of trustworthiness expectations in terms of the predicted behavior of others, Essay #2.2 focuses on the obligations which trustors assign to the trustee role (hereafter denoted as "trustworthiness obligations"), which can be considered as trustors' moral demands toward trustees and thus as a particular type of expectation. We find a negative correlation between trustworthiness obligations and one's willingness to trust. As an underlying reason, we suggest that individuals who have internalized social norms related to trustworthiness are wary of these norms being violated by others. This finding is somewhat similar to Essay #2.1, where we suggest that people seek to avoid being worse off than they initially expected. Technically, one might argue that trustworthiness obligations are positively

correlated with trustworthiness beliefs. In this case, the negative effect of trustworthiness obligations on trust would be only a variation of the findings reported in Essay #2.1. To invalidate this concern, we control for trustworthiness beliefs, measured in a similar manner to Essay #2.1, in the regression analysis of Essay #2.2. In this vein, we can show that the effect of trustworthiness obligations revealed by Essay #2.2 is empirically distinct from the effect of avoiding unfulfilled trustworthiness beliefs put forth in Essay #2.1.

Essay #2.3 studies the intrapersonal relationship between trust and trustworthiness by means of a within-subject experiment, in which participants take the trustor role and then take the trustee role, and vice versa. We find that trustworthiness beliefs serve as a moderator for the relationship between trust and trustworthiness within individuals (see arrow labeled "4"): Optimistic beliefs promote a positive relationship between trust and trustworthiness, whereas the contrary is true for pessimistic beliefs. It is obvious that Essay #2.3 is also closely related to Essays #2.1, because both essays study the relationship between social preferences, measured either in terms of social value orientation (Essay #2.1) or trustworthiness (Essay #2.3), and trustworthiness beliefs. However, the implications of both essays are complementary rather than interchangeable: As Figure 5 points out, Essay #2.1 stresses the role of trustworthiness beliefs as a *mediator* for the relationship between social preferences and trust, whereas Essay #2.3 suggests that trustworthiness beliefs serve as a *moderator*.

The similar scope of Essays #2.1 and #2.3 might raise the question of why two distinct measures of social preferences are used. In fact, it is plausible to assume that the strength of the observed effects depends on whether social preferences are measured in terms of either social value orientation or of trustworthiness, because both measures differ in a crucial aspect. The concept of social value orientation captures outcome-based preferences only (Murphy

& Ackermann, 2014) (see Section 3 for further details). Trustworthiness, by contrast, is a result of reciprocal preferences and incorporates thus both an outcome-based and an intentions-based preference component. This conceptual distinction, which is illustrated in Figure 6, is especially relevant for the moderator effect. As outlined in Section 4.1, the reason for the moderator effect is that individuals holding social preferences experience additional emotions when being rewarded or being betrayed, compared to fully selfish individuals: Trustworthy individuals experience additional positive emotions when being rewarded and additional negative emotions when being betrayed. Obviously, these emotions depend on how people evaluate the actions of others with respect to the latter's intentions. Therefore, the moderator effect is presumably more pronounced if social preferences are measured in terms of intensions-based preferences, i.e. in terms of trustworthiness. The mediator effect, by contrast, ought to depend to a smaller extent on the way social preferences are measured. As outlined earlier, the mediator effect mainly builds on a consensus effect, i.e. a strong connection between social preferences and trustworthiness beliefs. Intuitively, one would expect trustworthiness to be an even better predictor of trustworthiness beliefs than social value orientation is. However, a major advantage of using social value orientation as a measure of social preferences is that the experimental elicitation of trust and the experimental elicitation of social value orientation are contextually independent of each other.

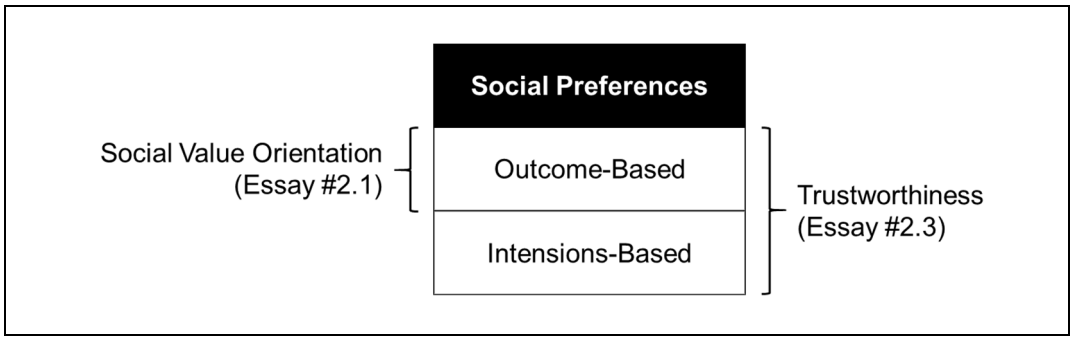

Figure 6: Measures of social preferences used in Essays #2.1 and #2.3

4.3 Overall Contribution, Limitations, and Open Questions

The contribution of the essays included in the present thesis is multi-dimensional. At the meta-level, each essay shows the "flip side" of a well-established behavioral mechanism. In the context of trust, an example of such a behavioral duality can be found in Buchan et al. (2008), who investigate whether feelings of obligation affect behavior in the trust game. The authors find that trustors who feel obligated to trust are more likely to send any money compared to trustors who lack such obligation feelings. This finding is not surprising in the light of the social dimension of trust outlined in Section 3. Female trustors, however, seem to behave in opposition to their obligation feelings: The degree to which they feel obligated to trust is negatively correlated with their amounts sent. Buchan et al. (2008) attribute this counterintuitive finding to a reactance effect (Brehm, 1966). According to this effect, the feeling of being obligated to take a certain action limits the range of choices, which people perceive as being available to them. This perceived limitation may motivate people to behave in opposition to their obligation.

The essays included in the present thesis underline the importance of dual relationships in two particular situations involving social risk: the two-player market entry game and the binary-choice trust game. The findings reported in the essays are similar to Buchan et al.

(2008), as they indirectly or directly reveal counteracting effects. Moreover, the authors point out that certain effects may have multiple causes. In the following, I briefly summarize the dualities revealed in this thesis:

- In Essay #1, I identify a beliefs-based and a preference-based relationship between one's behavior in a two-player market entry game and the physical attractiveness of one's opponent. Future research should examine whether this duality can be found in other economic contexts, where, for example, the two effects of appearance counteract each other.
- In Essay #2.1, we encounter dualities in two aspects. The first dual finding concerns the way in which beliefs are mapped into decisions. From the perspective of expected utility, one would expect optimistic trustworthiness beliefs to be associated with higher levels of trust (e.g. Berg et al., 1995). However, the effect of avoiding unfulfilled expectations examined in Essay #2.1 imposes a negative correlation between trustworthiness beliefs and trust. Future experimental research should assess the relative size of both effects. The negative effect of trustworthiness beliefs on trust induces a second duality: In analogy to trustworthiness beliefs, social value orientation carries out two counteracting effects on trust.
- Essay #2.2 reveals a duality with respect to norm orientation. We replicate the finding reported in Buchan et al. (2008) that feelings of obligation increase trustors' willingness to trust. However, we also observe that norm-orientated trustors who consider trustees to be obligated to reciprocate are less willing to trust than trustors who negate such an obligation.

– In Essay #2.3, we investigate the influence of trustworthiness on trust in a within-subject design. We find that trustworthiness carries out either a positive or a negative influence on trust, depending on whether people hold pessimistic or optimistic trustworthiness beliefs. The moderating effect of trustworthiness beliefs constitutes an indirect connection between trustworthiness beliefs and trust, which adds to the direct connections discussed in Essay #2.1.

The variety of dualities discussed in the essays underlines that understanding social risk taking behavior is challenging, since many of the underlying behavioral patterns are conditional. In this context, the present thesis especially emphasizes the relevance of moderator effects (Essays #1 and #2.3) and mediator effects (Essays #1 and #2.1), which both shape the triangular relationship between expectations, preferences, and social risk taking.

References

Aimone, J. A., & Houser, D. (2012). What You Don't Know Won't Hurt You: A Laboratory Analysis of Betrayal Aversion. *Experimental Economics*, *15*(4), 571–588.

Altmann, S., Dohmen, T., & Wibral, M. (2008). Do the Reciprocal Trust Less? *Economics Letters*, *99*(3), 454–457.

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

Arrow, K. J. (1981). Optimal and Voluntary Income Redistribution. In S. Rosenfield (Ed.), *Economic Welfare and the Economics of Soviet Socialism: Essays in Honor of Abram Bergson.* Cambridge: Cambridge University Press.

Aydogan, G. (2015). Trust and Social Risk: Neuro-Scientific Foundations of Economic Decision-Making Under Strategic Uncertainty (*Dissertation*). Ludwig-Maximilians-Universität, Munich.

Becker, G. S. (1974). A Theory of Social Interactions. *Journal of Political Economy*, *82*(6), 1063–1093.

Ben-Ner, A., & Putterman, L. (2001). Trusting and Trustworthiness. *Boston University Law Review*, *81*, 523–551.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *10*(1), 122–142.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2014). Preferences and Beliefs in a Sequential Social Dilemma: A Within-Subjects Analysis. *Games and Economic Behavior*, *87*, 122–135.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, *98*(1), 294–310.

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). Trust and the Reference Points for Trustworthiness in Gulf and Western Countries. *Quarterly Journal of Economics*, *125*(2), 811–828.

Bohnet, I., & Zeckhauser, R. (2004). Trust, Risk and Betrayal. *Journal of Economic Behavior and Organization*, *55*(4), 467–484.

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

Brehm, J. W. (1966). A Theory of Psychological Reactance. Oxford: Academic Press.

Breuer, W., Helduser, C., & Schade, P. (2016). Breaking the Rules: Anticipation of Norm Violation in a Binary-Choice Trust Game. *Economics Letters*, *146*, 123–125.

Bromiley, P., & Cummings, L. L. (1995). Transactions Costs in Organizations with Trust. In R. J. Lewicki & B. H. Sheppard (Eds.), *Research on Negotiation in Organizations* (Vol. 5, pp. 219–247). Greenwich: Jai Press.

Buchan, N. R., Croson, R. T. A., & Solnick, S. (2008). Trust and Gender: An Examination of Behavior and Beliefs in the Investment Game. *Journal of Economic Behavior and Organization*, *68*(3), 466–476.

Buchan, N. R., Johnson, E. J., & Croson, R. T. (2006). Let's Get Personal: An International Examination of the Influence of Communication, Culture and Social Distance on Other Regarding Preferences. *Journal of Economic Behavior and Organization*, *60*(3), 373–398.

Büyükboyacı, M. (2014). Risk Attitudes and the Stag-Hunt Game. *Economics Letters*, *124*(3), 323–325.

Charness, G. & Rabin, M. (2000). Social Preferences: Some Simple Tests and a New Model (*Unpublished Manuscript*).

Coleman, J. S. (1990). Foundations of Social Theory. Cambridge: Harvard University Press.

Corcos, A., Pannequin, F., & Bourgeois-Gironde, S. (2012). Is Trust an Ambiguous Rather Than a Risky Decision? *Economics Bulletin*, *32*(3), 2255–2266.

Costa-Gomes, M. A., Huck, S., & Weizsäcker, G. (2014). Beliefs and Actions in the Trust Game: Creating Instrumental Variables to Estimate the Causal Effect. *Games and Economic Behavior*, *88*, 298–309.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Dufwenberg, M., & Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Eckel, C. C., & Wilson, R. K. (2004). Is Trust a Risky Decision? *Journal of Economic Behavior and Organization*, *55*(4), 447–465.

Engel, C. (2011). Dictator Games: A Meta Study. *Experimental Economics*, *14*(4), 583–610.

Engelmann, D., & Strobel, M. (2004). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *American Economic Review*, *94*(4), 857–869.

Falk, A., & Fischbacher, U. (2006). A Theory of Reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Fehr, E., & Fischbacher, U. (2002). Why Social Preferences Matter: The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives. *Economic Journal*, *112*(478), C1-C33.

Fehr, E. & Schmidt, K. M. (2000). Theories of Fairness and Reciprocity: Evidence and Economic Applications (*Unpublished Manuscript*).

Fehr, E., & Schmidt, K. M. (2006). The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories. In S.-C. Kolm & J. M. Ythier (Eds.), *Handbooks in Economics: Vol. 23. Handbook of the Economics of Giving, Altruism and Reciprocity* (pp. 615–691). Amsterdam: North-Holland.

Fetchenhauer, D., & Dunning, D. (2009). Do People Trust Too Much or Too Little? *Journal of Economic Psychology*, *30*(3), 263–276.

Fetchenhauer, D., & Dunning, D. (2012). Betrayal Aversion versus Principled Trustfulness: How to Explain Risk Avoidance and Risky Choices in Trust Games. *Journal of Economic Behavior and Organization*, *81*(2), 534–541.

Griesinger, D. W., & Livingston, J. W. (1973). Toward a Model of Interpersonal Motivation in Experimental Games. *Systems Research and Behavioral Science*, *18*(3), 173–188.

Heinemann, F., Nagel, R., & Ockenfels, P. (2009). Measuring Strategic Uncertainty in Coordination Games. *Review of Economic Studies*, *76*(1), 181–221.

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, *92*(5), 1644–1655.

Houser, D., Schunk, D., & Winter, J. (2010). Distinguishing Trust from Risk: An Anatomy of the Investment Game. *Journal of Economic Behavior and Organization*, *74*(1), 72–81.

Johansson-Stenman, O., Mahmud, M., & Martinsson, P. (2005). Does Stake Size Matter in Trust Games? *Economics Letters*, *88*(3), 365–369.

Johnson, N. D., & Mislin, A. A. (2011). Trust Games: A Meta-Analysis. *Journal of Economic Psychology*, *32*(5), 865–889.

Kagel, J. H., & Roth, A. E. (Eds.). (2009). The Handbook of Experimental Economics. Princeton: Princeton University Press.

Kahneman, D. (1988). Experimental Economics: A Psychological Perspective. In R. Tietz, W. Albers, & R. Selten (Eds.), *Bounded Rational Behavior in Experimental Games and Markets* (pp. 11–18). Berlin: Springer.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *American Economic Review*, *76*(4), 728–741.

Knack, S., & Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, *112*(4), 1251–1288.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal Determinants of External Finance. *Journal of Finance*, *52*(3), 1131–1150.

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

Liebrand, W. B. G. (1984). The Effect of Social Motives, Communication and Group Size on Behaviour in an N-Person Multi-Stage Mixed-Motive Game. *European Journal of Social Psychology*, *14*(3), 239–264.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734.

McKnight, D. H., & Chervany, N. L. (2001). Trust and Distrust Definitions: One Bite at a Time. In R. Falcone, S. Munindar, & T. Yao-Hua (Eds.), *Trust in Cyber-Societies: Integrating the Human and Artificial Perspectives* (pp. 27–54). Berlin: Springer.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The False Consensus Effect: A Meta-Analysis of 115 Hypothesis Tests. *Journal of Experimental Social Psychology*, *21*(3), 262–283.

Murphy, R. O., & Ackermann, K. A. (2014). Social Value Orientation: Theoretical and Measurement Issues in the Study of Social Preferences. *Personality and Social Psychology Review*, *18*(1), 13–41.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring Social Value Orientation. *Judgment and Decision Making*, *6*(8), 771–781.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*, *83*(5), 1281–1302.

Ross, L., Greene, D., & House, P. (1977). The "False Consensus Effect": An Egocentric Bias in Social Perception and Attribution Processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301.

Samuelson, P. A. (1993). Altruism as a Problem Involving Group Versus Individual Selection in Economics and Biology. *American Economic Review*, *83*(2), 143–148.

Schechter, L. (2007). Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay. *Journal of Economic Behavior and Organization*, *62*(2), 272–292.

Selten, R., & Güth, W. (1982). Equilibrium Point Selection in a Class of Market Entry Games. In M. Deistler, E. Fürst, & G. Schwödiauer (Eds.), *Games, Economic Dynamics, and Time Series Analysis* (pp. 101–116). Heidelberg: Physica-Verlag HD.

Sen, A. (1995). Moral Codes and Economic Success. In C. S. Britten & A. Hamlin (Eds.), *Market Capitalism and Moral Values.* Aldershot: Edward Elgar.

Smith, A. (1982). The Theory of Moral Sentiments. Indianapolis: Liberty Fund (Original work published 1759).

Trautmann, S. T., & Vieider, F. M. (2012). Social Influences on Risk Attitudes: Applications in Economics. In S. Roeser, R. Hillerbrand, P. Sandin, & M. Peterson (Eds.), *Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk* (pp. 575–600). Berlin: Springer.

Zak, P. J., & Knack, S. (2001). Trust and Growth. *Economic Journal*, *111*(470), 295–321.

The Twofold Beauty Premium: Preference-Based and Beliefs-Based Behavior in an Anti-Coordination Game

Philipp Schade*

Abstract

I study the role of physical attractiveness and gender in a two-player market entry game with a safe option and a risky option. Participants are provided with photos of their co-players' faces before they make their decisions. I find that co-players' attractiveness affects individual decision making through two distinct transmission channels: a stereotype-based and a preference-based channel. Both are interacted with gender. Females expect other females who are rated as physically attractive to be less risk averse. As subjects behave rationally regarding their beliefs, good-looking females face fewer risky decisions in same-sex dyads. Males enter the market less often when facing a female co-player. However, males' motivation is preference-based, because their predictions of females' behavior are not affected by females' appearance. Thus, physically attractive females face fewer market entries in both same-sex and mixed-sex settings, but for different reasons. This confirms the existence of a twofold "beauty premium" in the labor market.

JEL Classification: C91, D81, D84, J16

Keywords: Anti-Coordination Game, Physical Attractiveness, Stereotypes, Strategic Risk

* Philipp Schade, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 94774, Fax: +49 241 80 92163, E-Mail: philipp.schade@bfw.rwth-aachen.de.

1 Introduction

In many social interactions, good-looking people seem to benefit from their appearance and the way others respond to it. The seminal study by Hamermesh and Biddle (1994) revealed a significant "beauty premium" that enables attractive people to earn more money than unattractive people. The authors found that workers who were rated by interviewers as "above average" in physical attractiveness earn about 10 to 15 % more than workers who were judged to be less physically attractive. Up to now, the existence of a "beauty premium" has been documented in various economic contexts, such as fund-raising (Landry, Lange, List, Price, & Rupp, 2006), peer-to-peer lending (Ravina, 2012), professional sports (Berri, Simmons, VanGilder, & O'Neill, 2011), and political elections (Berggren, Jordahl, & Poutvaara, 2016).

In the light of the significant economic impact of appearance, it is not surprising that the underlying behavioral mechanisms have attracted growing interest over the last years. The literature on physical attractiveness proposes two channels through which beauty may operate. According to the taste-based discrimination theory by Becker (1957), employers and customers derive utility ("pleasure") from interacting with physically attractive employees. Thus, employers pay them higher wages in order to extend the time span in which they remain together. Another strand of literature focusses on stereotypes related to physical attractiveness. Results from these studies indicate that people perceive beauty to be correlated with intelligence and mental health (Eagly, Ashmore, Makhijani, & Longo, 1991; Feingold, 1992; Jackson, Hunter, & Hodge, 1995; Kanazawa & Kovarb, 2004). Moreover, good-looking people are expected to behave more socially (Andreoni & Petrie, 2008; Kahn, Hottes, & Davis, 1971; Mulford, Orbell, Shatto, & Stockard, 1998; Solnick & Schweitzer, 1999; Wilson & Eckel, 2006). In a more recent

study considering an experimental labor market and a real-effort task for which physical attractiveness does not improve productivity, Mobius and Rosenblat (2006) found that employers expected physically attractive workers to perform better at their task. In summary, research on stereotypes related to appearance suggests that employers expect to derive positive *monetary* utility from favoring good-looking employees, contrasting with non-monetary utility, which is emphasized by preference-based approaches such as the seminal work of Becker (1957).

My study captures a key feature of competitive (labor) markets: the willingness to take *strategic* risk and to compete for a scarce resource which cannot be divided between different subjects, e.g. jobs and promotions.[1] To my knowledge, there is only one study focusing on the effect of beauty and gender matching in competitive environments. Dreber, Gerdes, and Gränsmark (2013) analyze the impact of appearance on strategies chosen in professional chess matches. They find that males but not females take more risk against attractive players of the opposite sex. However, this study differs from mine in two ways. On the one hand, it examines the risk-taking behavior of individuals who have already decided to take part in a competition and have self-selected into a high-risk environment. I look rather at the willingness to *enter* a competition and at the risk which is associated with this decision. On the other hand, chess players may choose certain strategies to prove their skills and to impress their opponents. In my experiment, such performance considerations are not a relevant factor.

To shed more light on the role of physical attractiveness in competitive environments, I study three research questions. First and foremost, I examine whether subjects' willingness to take strategic risk is affected by co-players' appearance. A two-step analysis of beliefs and

[1] For the sake of brevity, I do not distinguish between risk and ambiguity in this paper. In the following, I will speak of risk even if there is ambiguity.

actual decisions is conducted to show whether attractiveness affects decisions through a stereotype-based or a preference-based mechanism, or both. Since information on gender may facilitate coordination between two subjects (Holm, 2000), I investigate in a second step whether the potential effect of beauty is conditioned on gender-matching aspects. In the two-player market entry game which is studied in this paper, behavior should at least theoretically depend on how a subject expects to be perceived by the interaction partner. Therefore, as my third research question, I also examine whether risk-taking behavior depends on decision makers' own attractiveness and if so, whether decision makers' and co-players' attractiveness interact with each other.

My experiment is largely inspired by the market entry games proposed by Selten and Güth (1982) and Kahneman (1988), in which players have to decide simultaneously whether to enter an experimental market with a limited capacity. In order to keep the experimental framework as simple as possible, I consider a two-player market entry game with a market capacity of one. In the first part of the experiment, subjects are matched anonymously. In the second part, subjects are provided with photos of their co-players' faces before they make their decisions. I make use of a photo analysis application called *nFace* to compute individual attractiveness ratings based on a person's facial symmetry. Facial symmetry has been shown to be a reliable predictor of perceived attractiveness (Gangestad, Thornhill, & Yeo, 1994; Grammer & Thornhill, 1994; Perrett et al., 1999; Rhodes, Proffitt, Grady, & Sumich, 1998). Measuring facial symmetry by means of software tools has only recently received attention in economic research (see, e.g., Hoegele, Schmidt, & Torgler, 2016; Ravina, 2012).

The present research offers three features which lead to significant differences from previous studies which have also examined the role of appearance in strategic interactions. First,

the social optimum of the two-person market entry game is non-cooperative. Overall payoffs are maximized if one player enters the market while the other one stays out. As mentioned before, this feature covers real-life situations which do not allow for equal payoff distributions. Second, the participants of my experiment can choose a riskless option, bringing the experiment in line with traditional market entry games involving more than two players. If a player decides to stay out of the market, her or his payoff is inevitably zero, no matter which alternative the co-player chooses. This feature is crucial in terms of my research question: A safe option is required to measure individual willingness to take strategic risk. Third, the experiment allows an undistorted view of the role of strategic risk-taking behavior, since task performance and beliefs related to performance aspects do not need to be considered here.

Besides the broad literature on the "beauty premium" in the labor market, my study contributes to two more strands of research. First, I contribute to the literature on stereotypes related to risk-taking behavior. In this field, Ball, Eckel, and Heracleous (2010) report that physically attractive individuals are expected to take more financial risk than less attractive ones in a lottery selection task. I extend this research by examining whether an equivalent stereotype occurs in strategic contexts. Second, I contribute to the literature on the role of focal points in coordination and anti-coordination games. Both classes of games refer to situations where players face strategic risk due to the existence of multiple equilibria in pure strategies. In coordination games, it is mutually beneficial for players to choose similar strategies. In anti-coordination games, by contrast, players prefer choosing opposite strategies. In this context, Holm (2000) shows that gender may serve as a focal point when two subjects have to decide simultaneously on how to divide up a given amount of money between themselves. The present study extends this strand of literature by examining whether physical attractiveness

provides information which helps subjects to anti-coordinate their actions in the two-player market entry game.

In a nutshell, I find that in the two-person market entry game beauty affects behavior through both of the transmission channels which have been proposed in the literature. On the one hand, I identify a stereotype relating beauty to strategic risk-taking behavior. Females expect physically attractive female co-players to enter the market with a higher likelihood than less attractive female co-players, and they adapt their behavior by refraining from market entry more often. On the other hand, I find evidence for preference-based decision making in the case of males facing female co-players. More precisely, males enter the market less often when facing good-looking female co-players. In this regard, my results are in line with taste-based discrimination theory and match basic principles from evolutionary biology.

As outlined above, the role of gender and appearance has been studied extensively in the economic literature. The present study extends this literature by showing that gender and appearance also affect decision making in competitive environments. As I conduct a controlled laboratory experiment, I can reveal how subjects expect each other to behave and how subjects map these beliefs into their actual decisions. Thus, I am not only capable of identifying stereotypes and preferences as underlying causes for actual behavior but of disentangling them too. To my knowledge, the present study is the first one showing that in strategic interactions appearance can operate through both the stereotype-based and the preference-based transmission channel.

The remainder of this paper is organized as follows. Section 2 describes the experimental framework. In Section 3, hypotheses are derived from the related literature. Section 4 introduces the experimental design and reviews the procedures. Sections 5 and 6 present the results. Section 7 concludes.

2 Two-Player Market Entry Game

In the market entry game studied in this paper, two players have to decide simultaneously and without the other's knowledge whether to enter a given market which may be profitably exploited by only one of them. The market capacity and the number of potential entrants are exogenously determined and are common knowledge among players. Table 1 shows the respective payoffs in experimental currency units (ECUs). A player who decides to enter the market earns a positive payoff (3.5 ECUs) if the co-player does not enter. On the other hand, a player entering the market suffers a loss (-3.5 ECUs) if the co-player enters the market as well. If a player decides to stay out of the market, her or his payoff is zero, irrespective of the co-player's choice. As it is mutually beneficial for both players to take different actions, the two-player market entry game constitutes an anti-coordination game. Formally, there are two Nash equilibria in pure strategies, where one player enters the market while the other player stays out, and vice versa. Due to the existence of multiple pure strategy Nash equilibria, the game has a third Nash equilibrium in mixed strategies, where subjects randomly choose between entering the market and staying out, selecting each option with a probability of 1/2.

>>> Insert Table 1 about here <<<

The two-player market entry game differs from traditional market entry games in two aspects. On the one hand, I consider a group of only two potential entrants, whereas other studies typically involve 15 or more players (see, e.g., Kahneman, 1988; Selten & Güth, 1982). Second, the market capacity is one in all rounds whereas in previous experiments with more players, the market capacity differed from round to round. Both simplifications constitute a game which is similar to the game of chicken. In this two-player game proposed by Rapoport and Chammah (1966), two car drivers are assumed to face each other on a road that is too narrow for both cars to pass at the same time. The respective payoff matrix, containing arbitrarily set numerical payoffs, is given in Table 2. The players have to decide simultaneously whether to yield to the other player or to drive on. The best outcome for each player (1 ECU) is to drive on while the other yields. The worst outcome for both players (-10 ECUs) occurs if both decide not to yield to each other. A player suffers also a loss (-1 ECU) if she or he decides to yield but the co-player does not. Taken together, the game of chicken does not allow players to avoid strategic risk, because payoffs always depend on co-players' actions. In market entry games, however, staying out of the market yields neither a gain nor a loss, and constitutes thus a safe option. The existence of such a safe option is a required feature of my study, since I seek to examine individuals' willingness to take strategic risk.

>>> Insert Table 2 about here <<<

3 Hypotheses and Related Literature

3.1 Stereotype Hypothesis

In the two-player market entry game described in the preceding section, subjects have to choose between a risky option and a safe option. If a subject decides to enter the market, her or his payoff solely depends on whether the co-player enters the market as well. On the other hand, if a subject does not enter the market, her or his payoff is inevitably zero, irrespective of the co-player's decision. In the light of this safe option, subjects might expect co-players' risk preferences to be a relevant factor in the game. As photos are the only source of information available to subjects, they may use them to predict their co-players' behavior. Moreover, stereotypes associated with appearance are likely to be applied, because the great majority of participants do not know each other personally and lack therefore valuable personal information.

Research from different disciplines suggests the existence of a beauty stereotype regarding risk-taking behavior. A substantial psychological literature shows that people who are physically more attractive are perceived by others to be physically and mentally healthier. Cunningham (1986) found that men judge women with more attractive faces to be more fertile and be likely to experience fewer medical problems. Grammer and Thornhill (1994) documented that opposite-sex raters judge more facially attractive people to be healthier than less attractive people. Kalick, Zebrowitz, Langlois, and Johnson (1998) replicated these findings using both opposite-sex and same-sex raters. Singh (1993) found that not only are women with a lower waist-to-hip ratio assumed to be physically healthier, more fertile, and more attractive, but

women with lower waist-to-hip ratio report indeed fewer health problems and higher fertility. Feingold (1992) shows that the beauty stereotype applies analogously to mental health.

Why should *financial* risk taking be linked to physical traits such as beauty? I follow the argumentation of Ball et al. (2010), who combine explanations from different fields of research. Evolutionary biology suggests that both physical and psychological traits that enhance the likelihood of survival become more prevalent in future generations. Physical prowess and risk-taking behavior are closely related, because the physical prowess of hunters ought to ensure success when they take physical risks. Thus, risk-taking behavior is especially helpful for survival if it is combined with physical prowess. That is the reason why both traits are believed to co-evolve. According to psychologists, who tend to view risk taking as a personality trait, physical risk preferences should be related to preferences regarding other kinds of risk. Another argument put forward by Ball et al. refers to background risk theory. This theory assumes that individuals face a portfolio of different kinds of risk that are diversified within individuals. Physical prowess reduces physical risk, enabling individuals to take more risk in other domains. Various studies from the field of household finance confirm this theory empirically (Eeckhoudt, Gollier, & Schlesinger, 1996; Eeckhoudt & Kimball, 1992; Jianakoplos & Bernasek, 1998; Pratt & Zeckhauser, 1987; Sunden & Surette, 1998). From the arguments outlined above I derive my first hypothesis:

Hypothesis 1: *Physically attractive subjects are expected to enter the market more often than less attractive ones due to a stereotype relating physical attractiveness to strategic risk preferences. If subjects rely on their predictions of co-players' behavior, attractiveness should also affect their decisions.*

3.2 Gender-Pairing Hypothesis

In my experiment, subjects are shown photos of their co-players' faces in order to provide information on physical attractiveness. The information on gender that is collaterally transferred might moderate the relationship between beauty and belief formation that has been suggested in Section 3.1, because gender itself may be a strong source of stereotypes. A vast body of literature has documented gender differences in risk taking, suggesting that females take less risk than males do (see Byrnes, Miller, & Schafer, 1999 for an extended overview of gender differences in risk taking). More recently, researchers have begun to report gender stereotypes related to financial risk taking (Eckel & Grossman, 2002, 2008; Siegrist, Cvetkovich, & Gutscher, 2002). These studies confirm that females are expected to be more risk averse than males, although females' actual behavior does not always meet this expectation.

In the light of the gender stereotype related to risk-taking behavior, participants in my experiment may use co-players' sex as an additional source of information to predict co-players' behavior. Implications from psychological dual process theories suggest that the gender stereotype may even dominate a potential beauty stereotype, due to differences in the way both types of information are cognitively processed. According to the dual process model by Brewer (1988), impression formation always starts with an identification stage. At this stage, the stimulus person is automatically classified into categories based on salient characteristics such as gender, age, or skin color. Related psychological research on impression formation confirms that gender judgment takes place on a shallower level than judgments on less salient dimensions. Bower and Karlin (1974) conduct a recognition test in which subjects who had to judge first whether target persons were honest or likeable performed much better than subjects who had judged target persons only in terms of gender. In a reaction time study reported by Brewer

(1988), subjects were provided with stimulus photos of persons, which contained clear clues regarding gender and occupation. Subjects were shown two stimulus photos together and were asked to indicate whether the pictures were the same or different in terms of gender or occupations. The experiment revealed that the response time was lower when subjects were supposed to judge occupations and faced stimulus photos of different gender, but not vice versa. This finding provides support for the conjecture that information on gender is automatically processed, even if gender is not task-relevant.

The implication that information on gender may be processed with priority to information on attractiveness is important because dual process theory suggests that processing will be stopped once a "satisfactory resolution of the stimulus information" has been achieved (Brewer, 1988). Experimental economists provide indirect evidence for this theory. Subjects who have to forecast risky decisions of others ignore personal information on the latter's actual risk preferences in favor of gender stereotypes when gender information is provided first (Grossman, 2013). When gender information is provided second, personal information on risk preferences is taken into account. Thus, the dominance of gender information may also lead to a neglect of alternative information.

If the reasoning outlined above holds, the dominance of gender information should solely depend on whether subjects expect gender to be a relevant solution for the problem at hand. As the two-player market entry game represents an anti-coordination problem, I argue that any information only turns out to be helpful if the information-related stereotype expects players to act differently. Thus, the gender dominance hypothesis requires mixed-sex pairings. In same-sex pairings, the gender stereotype would predict both players to act in the same way and would not help subjects to solve the anti-coordination problem. Holm (2000) shows for an

experimental battle of the sexes game that gender pairing is indeed crucial for coordination and that mixed-sex pairings lead to an increase in earnings. Thus, the gender dominance hypothesis that I developed on the ground of psychological dual process theory has to be modified with respect to gender pairing. I propose the following hypothesis:

Hypothesis 2: *In mixed-sex settings, the effect of physical attractiveness on belief formation is smaller than in same-sex settings, or may even be absent, because the potential beauty stereotype is dominated by a gender stereotype.*

3.3 Preference Hypothesis

Several field studies have documented that males respond sensitively to female attractiveness in economic contexts. For instance, men's demand for consumer credits increases if a picture of an attractive female is included in the advertisement (Bertrand, Karlan, Mullainathan, Shafir, & Zinman, 2010). Female attractiveness produces more cooperative behavior by male counterparts in high-stakes television game shows (Darai & Grätz, 2013). Men's contribution to a charity increases when they face an attractive female fundraiser (Landry et al., 2006). All these findings reflect at least partly preference-based behavior, since they match basic principles from evolutionary biology, which assumes attractiveness to be a proxy for fertility. However, preference-based behavior with respect to appearance may be difficult to identify in anonymous environments, such as economic laboratory experiments, where participants are usually recruited from large subject pools. According to Becker (1957), taste-based discrimination should be most pronounced whenever a subject has a reasonable expectation that she or he will experience further interaction with the other person.

In fact, experimental research has observed taste-based discrimination under specific conditions only. Mobius and Rosenblat (2006) study a bargaining game where "employers" forecast productivity of "workers" who perform a real-effort task. The experiment includes two treatments which differ in the way employers' forecasts affect workers' wages. In the first treatment, workers' wages highly depend on employers' forecasts whereas in the second treatment, the relationship between forecasts and wages is set to "weak". A comparison of the two treatments does not provide evidence for taste-based discrimination, indicating that employers do not increase their forecasts of good-looking workers' productivity to raise workers' wages when forecasts are highly relevant for wages. Rosenblat (2008) conducts a dictator game experiment and finds that taste-based discrimination in favor of physically attractive subjects is only present when dictators are provided with both a photo and a recorded speech of the receiver. Photos alone do not induce subjects to be more altruistic toward good-looking co-players.

However, there are at least three arguments for why preference-based discrimination is a relevant factor in my experimental setting. First, the maximum opportunity costs of staying out of the market can be easily calculated by participants. Second, the option which favors the co-player is also the safe one. Subjects may use that as an internal justification for their behavior. In the labor market experiment of Mobius and Rosenblat (2006), employers would have been suffering from cognitive dissonance if they had favored good-looking workers because doing so would have required an intentional misjudgment of the latter's productivity. Third, the participants of my experiment are equally endowed with information. Each participant is shown a photo of her or his co-player, knowing that the co-player is being provided with a photo analogously. Moreover, after the conclusion of the experiment, all players' decisions are revealed to the respective co-players. Thus, responsibility of decisions can be clearly assigned

to a specific person. Individuals, who typically care about how they are perceived by others, may take this into consideration. In this regard, my study contrasts with the dictator game experiment of Rosenblat (2008) in which receivers do not get to know dictators at all. Taken together, I conclude:

Hypothesis 3: *Physically attractive subjects face fewer entries in the market entry game because their co-players want to do them a favor by staying out of the market. This effect especially applies to the case of males facing female co-players.*

4 Methods

4.1 Experimental Design

The experiment consists of three independent parts which are completed in the same order by all participants of the experiment. Subjects are told at the beginning of the experiment that they will be compensated for only one of the first two parts and that the part to be remunerated will be drawn by lot at the very end of the experiment. Throughout the entire experiment, payoffs are calculated in experimental currency units (ECUs). One ECU is equal to one Euro.

In the first part, subjects' non-strategic risk attitudes are elicited by means of a multiple price list containing nine rows (see Table 3). In the multiple price list, each row represents a decision problem including two options between which subjects have to choose, a certain payoff and a lottery. Whereas the lottery remains equal across rows, the certain payoff increases from row to row. Risk preferences can be compared between individuals by considering the number of lotteries chosen, under the assumption that subjects switch only once from the lottery to

the certain payoff. An increase in the number of chosen lotteries indicates a decrease in risk aversion. Before subjects answer the multiple price list, they are provided with further payoff information. More precisely, they are informed that if the first part of the experiment is chosen for payoff, only one of the nine decisions will be remunerated and that the decision to be remunerated will be chosen by lot at the end of the experiment.

>>> Insert Table 3 about here <<<

In the second part of the experiment, subjects take part in the two-player market entry game presented in Section 2. In total, subjects sequentially play five rounds of the game. To avoid learning effects within the market entry game and biased behavior in the subsequent part of the experiment, the outcomes of all five rounds (as mentioned above) are only disclosed to participants at the very end of the experiment. In each round, each subject is randomly paired with another one. Besides pairings, rounds differ in a second aspect. In the first round, subjects are not given any information about their co-players. They are only told that they will be randomly matched with another participant and that co-players' identity will not be revealed at any time of the experiment. This point is notable because subjects might wonder whether the photo taken at the beginning of the experiment might play a role in this part of the experiment or later. From the second to the fifth round, each subject is provided with a photo of her or his co-player's face before she or he makes a decision. The photos were taken before the beginning of the experiment. I conduct the first round under complete anonymity to elicit subjects' willingness to take strategic risk, irrespective of co-players' characteristics.

Apart from the varying amount of given information, each round of the market entry game follows the same procedure. First, subjects make their market entry decisions. Second,

they are asked to predict their co-players' decision and to state their confidence with respect to their prediction on a ten-point Likert scale. I elicit individual confidence levels in order to find out whether subjects expect co-players to apply mixed strategies, as described above. To be more precise, I believe that low confidence levels indicate an expectation of mixed strategies. I expect a weaker relationship between subjects' beliefs about their co-players' behavior and their own market entry decisions in the case of low confidence levels than in the case of high confidence levels.

If the second part of the experiment is chosen for payoff, only one of the five market entry decisions and only one of the five predictions are actually remunerated. Both decision and prediction are drawn by lot at the very end of the experiment. In the market entry game, subjects are paid according to the payoff matrix presented in Table 1. In order to avoid any net losses from the entry decision, each subject is initially endowed with 5.00 ECUs. A correct prediction of a co-player's decision yields 2.00 ECUs. In order to prevent subjects from hedging the risk associated with their market entry decision through their statement of beliefs, the drawing procedure does not allow decisions and predictions to be drawn from the same round. Subjects are informed about this constraint, without making explicit mention of the underlying reasons. In Section 5.1, I discuss in more detail why hedging does not seem to play a role in my experiment.

In the third part of the experiment, subjects are shown pictures of all other subjects' faces and are asked to rate their attractiveness on a ten-point Likert scale. Moreover, I ask subjects whether they know the person in the photo. Subjects are told that their ratings concerning co-players' attractiveness will not be disclosed among participants in order to allow an unbiased elicitation of preferences. In this vein, I obtain $n - 1$ attractiveness ratings per subject,

assuming that n subjects participate in each session. Subjects are not remunerated for their evaluation of co-players' attractiveness. Individual attractiveness ratings are considered in Section 6 to test the robustness of my results.

4.2 Procedures

The experiment was conducted in the experimental laboratory of a large German university, using z-Tree (Fischbacher, 2007). Subjects were recruited via ORSEE (Greiner, 2004). I conducted two sessions with 28 subjects per session, obtaining 56 participants in total. Overall, 30 female and 26 male subjects took part in the experiment. In both sessions, 15 females and 13 males showed up. The well-balanced sex-ratio in both sessions might have prevented subjects from expecting gender composition to be a relevant factor in the first round of the market entry game.

Upon arrival, participants were informed that a digital photo of their face would be taken before the beginning of the experiment and that their photo would be shown to other participants in the following experiment. Further details of the experiment were not disclosed. Moreover, participants were told that the photos would be used for research purpose only. All subjects accepted these conditions and had their portrait taken in a separated part of the laboratory, in front of a plain-colored wall. Before the photos were actually taken, subjects were asked to look straight into the camera with a neutral facial expression and a closed mouth. Photos were re-taken if subjects smiled or kept their eyes shut.

At the beginning of the experiment, subjects were provided with instructions covering the first part (elicitation of risk preferences) and the first round of the second part (market entry game). Subjects were told that the first part of the experiment was to be kept anonymous.

The instructions for the subsequent four rounds were distributed after subjects had completed the first round. Subjects were informed that the decisions they would make in the four subsequent rounds would be revealed to their counterparts at the end of the experiment. Sessions lasted about 45 minutes. Payments were made in cash in a separate room immediately after the experiment. Anonymity among subjects was ensured during the payment procedure.

5 Results

5.1 Descriptive Statistics

In the first part of the experiment, subjects' non-strategic risk preferences were elicited by means of a multiple price list. On average, subjects chose the lottery 3.70 times before switching to the certain payoff (see Table 4). This number indicates slight risk aversion, since the certain payoff equals the expected value of the lottery only in the fifth row of the multiple price list (see Table 4). I do not encounter any subject who switched from the lottery to the certain payoff, and vice versa, more than once. An independent-samples *t*-test is conducted to compare the number of chosen lotteries for females ($M = 3.67$, $SD = 1.63$) and males ($M = 3.73$, $SD = 1.29$). The test does not reveal statistical differences in risk-taking behavior between females and males ($t(54) = 0.56$, $p = 0.580$).

>>> Insert Table 4 about here <<<

In the second part of the experiment, subjects played five subsequent rounds of the market entry game. In the first round, 53.6 % of all subjects predicted that the anonymous co-player would enter the market (see Table 5). The actual market entry rate was 67.9 %. Thus,

subjects tended to underestimate the actual market entry rate on average. However, a McNemar test does not suggest a statistically significant difference between predictions and actual decisions ($\chi^2(1,\ N = 56) = 2.29,\ p = 0.190$). On average, females expected a higher number of market entries than males did in the anonymous round. According to a Fisher's exact test, this difference is not statistically significant ($p = 0.421$). Compared to the first round, both the predicted market entry rate and the actual market entry rate decreased in the subsequent four rounds. The predicted market entry rate decreased from 53.6 to 47.8 %, whereas the actual market entry rate fell from 67.9 to 58.8 %. Additional McNemar tests suggest that only the decrease of the rate of actual decisions is statistically significant ($\chi^2(1,\ N = 224) = 5.38,\ p < 0.05$).

>>> Insert Table 5 about here <<<

The experiment was designed in a manner that would prevent subjects from hedging the risk associated with their market entry decisions through their statement of beliefs. Note that the market entry decision and the prediction, which are remunerated if the second part of the experiment is chosen for payoff, are not drawn from the same round. However, it can be argued that subjects still have the possibility to hedge by choosing the same option in the decision task (e.g. to enter the market) and the "opposite" option (e.g. to predict co-player to enter the market) in the prediction task in each round. I believe that this is not an issue, because I do not encounter any subject who did not vary her or his behavior in the market entry game, either in the decision or in the prediction task.

As mentioned in the introduction, subjects' facial attractiveness scores are obtained from an application for mobile devices (*nFace*), which computes a facial beauty score on a scale

between 0 (lowest) and 10 (highest), based on a person's facial geometry. In this manner, basic insights from psychology which emphasize the positive relationship between facial symmetry and perceived attractiveness of individuals are taken into account. According to an independent samples t-test, attractiveness scores computed by *nFace* do not significantly differ ($t(54) = 0.053$, $p = 0.958$) between females ($M = 7.027$, $SD = 1.079$) and males ($M = 7.192$, $SD = 0.911$).

5.2 Stereotype-Based Decision Making

In this section, I discuss whether decision making in the two-player entry game is driven by beauty stereotypes, as pointed out in Section 3.1. Formally speaking, I examine whether co-players' appearance affects subjects' decision making *indirectly* through the mediation of beliefs. The analysis of the potential indirect effect consists of two steps. First, I investigate whether co-players' appearance affects subjects at the level of beliefs formation (Condition 1). Second, I examine whether beliefs influence subjects at the level of actual decision making when controlling for co-players' appearance (Condition 2). Note that the two conditions are derived from the causal steps method developed by Baron and Kenny (1986). If both conditions are fulfilled, the statistical significance of the indirect effect is tested.

As outlined in Section 3.1, I hypothesize that subjects form their beliefs based on beauty stereotypes related to risk preferences, expecting attractive subjects to enter the market with a higher probability (Hypothesis 1). According to Hypothesis 2, this effect may interact with gender because gender itself may be an important source of stereotypes in the context of the market entry game. More precisely, gender may serve as a "first-order" stereotype which overlays the attractiveness stereotype. Thus, I expect the effect of attractiveness to be dampened

or even not to be observable when gender enables anti-coordination in the market entry game. Specifically, I hypothesize that this condition is satisfied in mixed-sex settings, where females are expected to stay out of the market while males are expected to enter.

To study beliefs formation, I run probit regressions with subjects' predictions of co-players' decision as dependent variable (see Table 6). I include decision makers' attractiveness and co-players' attractiveness as explanatory variables in the regressions. Decision makers' attractiveness is taken into account to examine whether subjects expect appearance to be meaningful in terms of a coordination device or whether they use it as a proxy for others' preferences. To identify potential effects of gender matching (Hypothesis 2), I run regressions for the whole sample as well as for nine different subsamples, varying decision makers' sex only (see Regressions 2 and 3 in Table 6), co-players' sex only (Regressions 4 and 5), and both parties' sex (Regressions 6 to 9). In Regression 10, I consider same-sex dyads only.

>>> Insert Table 6 about here <<<

By including subjects' predictions of anonymous co-players' behavior in the models, I account for subjects' beliefs concerning others' strategic risk preferences, irrespective of individual characteristics. Furthermore, I control for decision makers' financial risk attitude (derived from the multiple price list) and for the age of decision makers and co-players. Age may be a driver for both expectations and decision making, because according to dual process theory presented in Section 3.2, age may help subjects to categorize others at the first stage of judgment. In this aspect, age may exercise the same function as gender. Moreover, age may be related to the perceived attractiveness of others. However, I do not expect age to be crucial in my experiment, because participants were relatively homogeneous in terms of their age.

According to Hypotheses 1 and 2, I expect co-players' attractiveness to be relevant in the cases of same-sex matching, i.e. in Regressions 6, 9, and 10. Regression analysis provides partial support for this conjecture: The coefficient of co-players' attractiveness is statistically significant in the case of females facing female co-players only (see Regression 6). The direction of the effect indicates that females expect physically attractive females to enter the market with a higher likelihood. The pseudo *R*-squared of Regression 6 is the highest one compared to the other regressions, which underlines the explanatory power of co-players' appearance in this case. Thus, Condition 1 is satisfied for the case of females facing female co-players. Note that the models reported in Table 6 do not reveal any significant predictors other than the variables related to attractiveness, which also illustrates the importance of attractiveness at the level of beliefs formation.

As outlined above, an indirect effect of co-players' appearance additionally requires beliefs to be predictive for market entry decisions when controlling for co-players' appearance (Condition 2). To test whether this condition is fulfilled, I perform another set of probit regressions for the whole sample and different subsamples (see Table 7). In each model, the actual market entry decision is considered as the independent variable. In addition to co-players' appearance, I include decision makers' financial risk attitude, their decisions made in the anonymous round, their stated confidence, their own attractiveness, and the age of decision makers and co-players as explanatory variables. For each model presented in Table 7, including the case of females facing female co-players, the effect of "prediction" on market entry decisions is statistically significant: The expectation that the co-player will enter the market lowers

subjects' willingness to enter the market, and vice versa.[2] Thus, both conditions indicating an indirect effect of co-players' appearance on market entry decisions are satisfied for the case of females facing female co-players.

>>> Insert Table 7 about here <<<

Finally, I examine whether the indirect effect is statistically significant. In linear regression models, the coefficient of an indirect effect can be computed by multiplying the coefficients of the partial effects. Since I carry out non-linear regression models, the coefficients of the partial effects are in different scales and have to be standardized before multiplying them (Kenny, 2013; MacKinnon & Dwyer, 1993; Singh, 1993). The significance of the indirect effect is tested by means of non-parametric bootstrapping using 10,000 resamples (Preacher & Hayes, 2004). The 95 % bias-corrected confidence interval for the indirect effect ranges from -0.40 to -0.06. As this interval does not include zero, the indirect effect is statistically significant. Thus, Hypothesis 1 cannot be rejected for the case of females facing female co-players. As these findings do not apply to mixed-sex dyads, Hypothesis 2, which suggests a difference between same-sex and mixed-sex dyads, is partially confirmed.

Up to this point, I have found that subjects are affected by co-players' appearance through stereotypes under specific conditions. As outlined earlier in this paper, subjects may also expect their own appearance to affect co-players' behavior, via stereotypes or appearance-related preferences. Previous regressions suggest that this does apply to females who face male

[2] This finding is notable, because subjects seem to behave in a "rational" manner, which may be an indicator for the application of pure strategies. If subjects applied mixed strategies, beliefs would not be relevant for decision making.

co-players (see Regression 7 in Table 6): Females rated as good-looking expect male counterparts to enter the market with a lower likelihood. Bootstrapping using 10,000 resamples indicates that the indirect effect of females' own appearance on their market entry decisions is statistically significant (bias-correcting bootstrapping 95 % confidence interval [0.03, 0.36]). Thus, females map their stereotype-based beliefs into actual decisions. I cannot clearly disentangle whether females' expectations refer to stereotype-based or preference-based behavior of males. However, my analysis shows that males' beliefs do not rely on beauty stereotypes at all (see Regressions 8 and 9 in Table 6). Interestingly, females do not expect their own appearance to affect other females. Hence, females stereotype each other based on appearance, but they do not correctly anticipate being stereotyped.

To complete my analysis of stereotype-based decision making, I study potential interactions between subjects' own appearance and co-players' appearance at the level of beliefs formation. For this purpose, I compute the difference between decision makers' and co-players' attractiveness score for each pair and include it as additional explanatory variable in the regressions. The respective results presented in Table 8 suggest that the difference of attractiveness scores does not affect beliefs formation. Apparently, subjects do not make use of appearance as a coordination device, as they do not rely on co-players' appearance in relation to their own appearance, and vice versa.

>>> Insert Table 8 about here <<<

In the present section, I have focused on the question of whether decision making in the two-player market entry game is affected by beauty stereotypes related to risk preferences. My

analysis provides partial support for Hypotheses 1 and 2. Findings indicating a beauty stereotype are limited to the case of females facing female co-players. Corresponding to Hypothesis 1, I observe that females expect good-looking females to enter the market with a higher probability. Additional statistical tests suggest that the beauty stereotype indirectly affects decision making through beliefs. Moreover, my analysis of beliefs formation reveals that females do not expect their own appearance to affect female co-players but do expect it to affect male ones.

5.3 Preference-Based Decision Making

In this section, I examine whether decision making in the two-player market entry is driven by preferences related to others' appearance and whether potential preference-based behavior interacts with subjects' own appearance. To this end, I first refer to the regressions shown in Table 7, where the actual market entry decision is considered as the dependent variable. Since I intend to identify a direct effect of appearance on decision making which is not mediated by players' beliefs, it is important that beliefs are controlled for.

Regression analysis indicates that co-players' attractiveness is relevant in the case of males facing female co-players (see Regression 7 in Table 7), which is in line with Hypothesis 3: Males refrain from entering the market when facing good-looking female co-players. Interestingly, this finding coincides with females' expectations, which have been discussed in the previous section. Good-looking females correctly anticipate that they will face fewer entries from males than less attractive females do (see Regression 7 in Table 6). Note that in contrast to co-players' appearance, subjects' own appearance does not influence decision making in any of the subsamples when controlling for predictions.

Corresponding to my previous analysis of beliefs formation, I run another set of regressions to identify potential interaction effects between subjects' and co-players' attractiveness score. Again, I include the difference of attractiveness scores instead of the two separate attractiveness variables. The results presented in Table 9 suggest that decision makers' and co-players' appearance interact only with each other in the case of males facing female co-players (see Regression 8). According to the previous finding of males favoring good-looking females over less attractive ones, I observe that the greater the difference of attractiveness, the higher is the likelihood of a market entry.

>>> Insert Table 9 about here <<<

In summary, the current section shows that co-players' appearance affects decision making through a preference-based channel in the case of males facing female co-players. Female attractiveness lowers male willingness to enter the market. As I control for beliefs (revealing that male decision makers do not expect attractive female co-players ceteris paribus to enter more often), I can clearly identify preferences as the underlying cause. Moreover, I find that decision makers' and co-players' attractiveness interact with each other in the case of males facing female co-players. The lower co-players' attractiveness score compared to males' own attractiveness score, the higher is the probability that the latter will enter the market.

6 Robustness Checks

To test the robustness of the previous results, I reproduce my analyses using an alternative measure of attractiveness. Instead of the attractiveness ratings computed by *nFace* (*symmetry*

measure), I consider the individual attractiveness ratings obtained from the third part of the experiment (*survey measure*). To ensure comparability between subjects in terms of their attractiveness ratings given for others, the latter are first standardized within subjects. Second, the mean of all ratings which a participant has received from other participants is computed. In order to obtain unbiased judgments, I exclude ratings from further analysis if raters have stated that they knew the target person. Females ($M = 0.21$, $SD = 0.72$) got higher ratings than males ($M =$ -0.24, $SD = 0.51$) did (see Table 4). An independent samples t-test shows that the difference between females and males in attractiveness scores is statistically significant ($t(54) = 2.487$, $p < 0.05$). In line with the literature, the facial symmetry scores obtained from *nFace* on the one hand and the individual attractiveness ratings on the other hand are moderately correlated on the overall level ($r = 0.309$, $p < 0.05$). Rhodes et al. (1998) report correlation coefficients between 0.27 and 0.43. In contrast to Rhodes et al. (1998), correlations differ substantially between females ($r = 0.460$, $p < 0.05$) and males ($r = 0.063$, $p = 0.761$) in my sample. If I control for age, the correlation coefficient for males increases to 0.219. However, the correlation is not statistically significant ($p = 0.292$). As my main results refer solely to female appearance, I do not expect this artefact to be crucial for further analysis.

Corresponding to Section 5, I first focus on stereotype-based behavior. Regression 6 in Table 10 confirms that females expect good-looking female co-players to enter the market with a higher likelihood compared to less attractive females. Regression 6 in Table 11 suggests that there is also a statistically significant negative relationship between beliefs and decisions when controlling for co-players' appearance. Thus, Conditions 1 and 2 are both satisfied. According to a bootstrapping procedure using 10,000 resamples, the indirect effect of co-players' appearance on market entry decisions is statistically significant (bias-correcting bootstrapping 95 %

confidence interval [-0.34, -0.03]). In addition to the indirect effect, I have previously found that females expect their own appearance to affect male co-players' behavior. However, I cannot reproduce this finding when using the survey measure (see Regression 7 in Table 10). Moreover, I do not observe any interaction effects with regard to attractiveness at the level of beliefs formation (see Table 12).

>>> Insert Tables 10, 11, and 12 about here <<<

My previous analysis suggests preference-based behavior in the case of males facing female co-players. This main finding is confirmed by Regression 8 in Table 11. The respective regression also reveals an additional result which has not been observed in the previous analysis: Good-looking males avoid market entry when the co-player is female. Therefore, the interaction effect reported in Section 5 disappears (see Regressions 8 in Tables 9 and 13). This finding may be interpreted in terms of taste-based discrimination theory: Good-looking males may be aware of their appearance and may have experienced their appearance being rewarded in social interactions. Put another way, good-looking males may expect to derive a higher "social" utility from staying out of the market than others. Although this finding seems to be in line with intuition and theory, it is not robust, since it cannot be observed when using the symmetry measure. One may speculate that this lack of robustness may be related to the inconsistency of attractiveness measures for males reported above.

>>> Insert Table 13 about here <<<

The present section is intended to test the robustness of my results by means of an alternative attractiveness measure. Reproducing my analyses from Section 5 provides further

evidence for the main results. On the one hand, I identify stereotype-based decision making in female dyads. Females expect good-looking females to take more strategic risk and the former behave rationally with regard to this belief. On the other hand, I observe preference-based behavior in the case of males facing female co-players: Males demonstrate favor to good-looking females by not entering the market. The finding that females correctly anticipate the preference-based behavior of males cannot be confirmed.

7 Conclusion

In this paper, I report results from a two-person market entry game with disclosed appearances to examine the relevance of attractiveness and gender for the willingness to take strategic risk. I find that appearance operates through both of the transmission channels which have been previously proposed in the literature. I observe both stereotype-based and preference-based behavior, depending on how sexes are matched. At the level of beliefs formation, I observe a stereotype that relates beauty to strategic risk-taking behavior. Female subjects expect good-looking females to enter the market more often than less attractive females. I argue that this finding is due to the fact that information on gender does not help subjects to coordinate in same-sex settings.

Actual market entry decisions are affected by appearance in two ways. As subjects strongly rely on their beliefs when making decisions, the beauty effect is also observable on the level of decisions, but limited to the case of females facing female co-players. Moreover, I provide evidence for preference-based behavior in the case of males facing female co-players. In line with taste-based discrimination theory and basic implications from evolutionary biology,

men seem to demonstrate favor to physically attractive women by refraining from market entry. I speculate that males intend to prevent good-looking females from monetary loss that would occur in the case of a simultaneous market entry. On the other hand, males might want to signal to females that they are not taking advantage of the circumstances that favor males in a mixed-sex dyad. A third result is that subjects take into account their own appearance only to a limited extent. Females rated as good-looking expect their appearance to lower males' willingness to enter the market, thereby correctly anticipating males' behavior. Note that this finding disappears when attractiveness is not determined according to facial symmetry but according to the assessment of all participants in the experiment. In female dyads, subjects' own appearance does not affect beliefs. This finding suggests that subjects do not use appearance as a coordination device.

The finding that the effects of attractiveness interact with gender corresponds to previous studies suggesting that men change their behavior when interacting with physically attractive women (e.g. Bertrand et al., 2010; Darai & Grätz, 2013; Dreber et al., 2013; Landry et al., 2006). However, the present study is the first one which explicitly shows that the interaction between appearance and behavior may take place at both levels – the level of beliefs formation and the level of actual decision making – and that the interaction varies with respect to these levels. Moreover, I demonstrate that the beauty stereotype observed in non-strategic contexts may also apply to strategic interactions, especially to competitive environments. Besides these contributions, my research raises questions that deserve further investigation. Most importantly, future research should address the question of why subjects who rely on a certain stereotype when forming beliefs do not anticipate that their counterparts' beliefs might be affected by the same stereotype.

As outlined at the very beginning, the findings from my study can be transferred to real world situations in which individuals have to decide whether to compete for a scarce resource that cannot be shared equally among them. Moreover, it is required that the bilateral decision to compete – the failure of coordination – causes harm to competitors. In entrepreneurship, firms often face the problem of whether to enter a new market. Under the condition of limited demand, excessive market entries may cause price wars that finally ruin all entrants. However, my research focusses rather on individuals than on organizations. Therefore, the implications from my study are probably better suited to the labor market, where employees have to decide frequently whether to compete with others, e.g. for a leadership position. Competition may especially harm employees if the expected gains from winning the competition are low. For instance, competition may cause disutility to competitors if a third party takes advantage of the situation. This phenomenon is common in the field of politics, e.g. when the opposition gains popularity just because the governing parties are involved in ongoing conflicts with each other.

Acknowledgments

I thank Wolfgang Breuer, Stefan Kleeschulte, Christine Stibbe, participants in the 2015 annual meeting of the German Association for Experimental Economic Research (GfeW) in Hamburg, seminar participants at the Management Academy of the German Savings Banks Association in Bonn, as well as participants in the MOE Doctoral Workshop at RWTH Aachen University for valuable comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Andreoni, J., & Petrie, R. (2008). Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments. *Journal of Economic Psychology*, *29*(1), 73–93.

Ball, S., Eckel, C. C., & Heracleous, M. (2010). Risk Aversion and Physical Prowess: Prediction, Choice and Bias. *Journal of Risk and Uncertainty*, *41*(3), 167–193.

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Becker, G. S. (1957). The Economics of Discrimination. Chicago: University of Chicago Press.

Berggren, N., Jordahl, H., & Poutvaara, P. (2016). The Right Look: Conservative Politicians Look Better and Voters Reward It. *Journal of Public Economics*,

Berri, D. J., Simmons, R., VanGilder, J., & O'Neill, L. (2011). What Does It Mean to Find the Face of the Franchise? Physical Attractiveness and the Evaluation of Athletic Performance. *Economics Letters*, *111*(3), 200–202.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment. *Quarterly Journal of Economics*, *125*(1), 263–305.

Bower, G. H., & Karlin, M. B. (1974). Depth of Processing Pictures of Faces and Recognition Memory. *Journal of Experimental Psychology*, *103*(4), 751–757.

Brewer, M. B. (1988). A Dual Process Model of Impression Formation. In R. S. Wyer & T. K. Scrull (Eds.), *Advances in Social Cognition. A Dual Process Model of Impression Formation* (Vol. 1). Hillsdale: Erlbaum.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender Differences in Risk Taking: A Meta-Analysis. *Psychological Bulletin*, *125*(3), 367–383.

Cunningham, M. R. (1986). Measuring the Physical in Physical Attractiveness: Quasi-Experiments on the Sociobiology of Female Facial Beauty. *Journal of Personality and Social Psychology*, *50*(5), 925–935.

Darai, D. & Grätz, S. (2013). Attraction and Cooperative Behavior (*Unpublished Manuscript*).

Dreber, A., Gerdes, C., & Gränsmark, P. (2013). Beauty Queens and Battling Knights: Risk Taking and Attractiveness in Chess. *Journal of Economic Behavior and Organization*, *90*, 1–18.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is Beautiful is Good, but...: A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype. *Psychological Bulletin*, *110*(1), 109–128.

Eckel, C. C., & Grossman, P. J. (2002). Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. *Evolution and Human Behavior*, *23*(4), 281–295.

Eckel, C. C., & Grossman, P. J. (2008). Men, Women and Risk Aversion: Experimental Evidence. In C. R. Plott & V. L. Smith (Eds.), *Handbook of Experimental Economics Results* (1st ed., Vol. 1, pp. 1061–1073). Amsterdam: North-Holland.

Eeckhoudt, L., Gollier, C., & Schlesinger, H. (1996). Changes in Background Risk and Risk Taking Behavior. *Econometrica*, *64*(3), 683.

Eeckhoudt, L., & Kimball, M. (1992). Background Risk, Prudence, and the Demand for Insurance. In G. Dionne (Ed.), *Contributions to Insurance Economics* (Vol. 13, pp. 239–254). Berlin: Springer.

Feingold, A. (1992). Good-Looking People are not What We Think. *Psychological Bulletin*, *111*(2), 304–341.

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

Gangestad, S. W., Thornhill, R., & Yeo, R. A. (1994). Facial Attractiveness, Developmental Stability, and Fluctuating Asymmetry. *Ethology and Sociobiology*, *15*(2), 73–85.

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) Facial Attractiveness and Sexual Selection: The Role of Symmetry and Averageness. *Journal of Comparative Psychology*, *108*(3), 233–242.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In V. Macho & Kremer K. (Eds.), *Forschung und wissenschaftliches Rechnen* (63rd ed., pp. 79–93). Göttingen: Gesellschaft für wissenschaftliche Datenverarbeitung.

Grossman, P. J. (2013). Holding Fast: The Persistence and Dominance of Gender Stereotypes. *Economic Inquiry*, *51*(1), 747–763.

Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the Labor Market. *American Economic Review*, *84*(5), 1174–1194.

Hoegele, D., Schmidt, S. L., & Torgler, B. (2016). The Importance of Key Celebrity Characteristics for Customer Segmentation by Age and Gender: Does Beauty Matter in Professional Football? *Review of Managerial Science*, *10*(3), 601–627.

Holm, H. J. (2000). Gender-Based Focal Points. *Games and Economic Behavior*, *32*(2), 292–314.

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review. *Social Psychology Quarterly*, *58*(2), 108.

Jianakoplos, N. A., & Bernasek, A. (1998). Are Women More Risk Averse? *Economic Inquiry*, *36*(4), 620–630.

Kahn, A., Hottes, J., & Davis, W. L. (1971). Cooperation and Optimal Responding in the Prisoner's Dilemma Game: Effects of Sex and Physical Attractiveness. *Journal of Personality and Social Psychology*, *17*(3), 267–279.

Kahneman, D. (1988). Experimental Economics: A Psychological Perspective. In R. Tietz, W. Albers, & R. Selten (Eds.), *Bounded Rational Behavior in Experimental Games and Markets* (pp. 11–18). Berlin: Springer.

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., & Johnson, R. M. (1998). Does Human Facial Attractiveness Honestly Advertise Health? Longitudinal Data on an Evolutionary Question. *Psychological Science*, *9*(1), 8–13.

Kanazawa, S., & Kovarb, J. L. (2004). Why Beautiful People are More Intelligent. *Intelligence*, *32*(3), 227–243.

Kenny, D. A. (2013). Mediation with Dichotomous Outcomes. Retrieved from http://davidakenny.net/doc/dichmed.pdf

Landry, C. E., Lange, A., List, J. A., Price, M. K., & Rupp, N. G. (2006). Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment. *Quarterly Journal of Economics*, *121*(2), 747–782.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating Mediated Effects in Prevention Studies. *Evaluation Review*, *17*(2), 144–158.

Mobius, M. M., & Rosenblat, T. S. (2006). Why Beauty Matters. *American Economic Review*, *96*(1), 222–235.

Mulford, M., Orbell, J., Shatto, C., & Stockard, J. (1998). Physical Attractiveness, Opportunity, and Success in Everyday Exchange. *American Journal of Sociology*, *103*(6), 1565–1592.

Perrett, D. I., Burt, D., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and Human Facial Attractiveness. *Evolution and Human Behavior*, *20*(5), 295–307.

Pratt, J. W., & Zeckhauser, R. J. (1987). Proper Risk Aversion. *Econometrica*, *55*(1), 143.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models. *Behavior Research Methods, Instruments, and Computers*, *36*(4), 717–731.

Rapoport, A., & Chammah, A. M. (1966). The Game of Chicken. *American Behavioral Scientist*, *10*(3), 10–28.

Ravina, E. (2012). Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets (*Unpublished Manuscript*).

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial Symmetry and the Perception of Beauty. *Psychonomic Bulletin and Review*, *5*(4), 659–669.

Rosenblat, T. S. (2008). The Beauty Premium: Physical Attractiveness and Gender in Dictator Games. *Negotiation Journal*, *24*(4), 465–481.

Selten, R., & Güth, W. (1982). Equilibrium Point Selection in a Class of Market Entry Games. In M. Deistler, E. Fürst, & G. Schwödiauer (Eds.), *Games, Economic Dynamics, and Time Series Analysis* (pp. 101–116). Heidelberg: Physica-Verlag HD.

Siegrist, M., Cvetkovich, G., & Gutscher, H. (2002). Risk Preference Predictions and Gender Stereotypes. *Organizational Behavior and Human Decision Processes*, *87*(1), 91–102.

Singh, D. (1993). Adaptive Significance of Female Physical Attractiveness: Role of Waist-to-Hip Ratio. *Journal of Personality and Social Psychology*, *65*(2), 293–307.

Solnick, & Schweitzer. (1999). The Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions. *Organizational Behavior and Human Decision Processes*, *79*(3), 199–215.

Sunden, A. E., & Surette, B. J. (1998). Gender Differences in the Allocation of Assets in Retirement Savings Plans. *American Economic Review*, *88*(2), 207–211.

Wilson, R. K., & Eckel, C. C. (2006). Judging a Book by its Cover: Beauty and Expectations in the Trust Game. *Political Research Quarterly*, *59*(2), 189–202.

Tables

		Player B	
		Stay Out	Enter
Player A	Stay Out	(0, 0)	(0, 3.5)
	Enter	(3.5, 0)	(-3.5, -3.5)

Table 1: Two-player market entry game

		Player B	
		Yield	Drive On
Player A	Yield	(0, 0)	(-1, 1)
	Drive On	(1, -1)	(-10, -10)

Table 2: Game of chicken

Decision	*Option A*	*Option B*	*Your Choice*
1	win 5.00 or nothing	get 0.50 for sure	A or B
2	win 5.00 or nothing	get 1.00 for sure	A or B
3	win 5.00 or nothing	get 1.50 for sure	A or B
4	win 5.00 or nothing	get 2.00 for sure	A or B
5	win 5.00 or nothing	get 2.50 for sure	A or B
6	win 5.00 or nothing	get 3.00 for sure	A or B
7	win 5.00 or nothing	get 3.50 for sure	A or B
8	win 5.00 or nothing	get 4.00 for sure	A or B
9	win 5.00 or nothing	get 4.50 for sure	A or B

Table 3: Multiple price list

Variable	*M*	Lowest	Highest	*SD*
Age	24.70	19	45	4.30
of Females	24.07	19	34	3.40
of Males	25.42	19	45	5.07
Number of Lotteries Chosen	3.70	0	7	1.48
by Females	3.67	0	7	1.63
by Males	3.73	0	7	1.30
Attractiveness (Symmetry Measure)	7.01	4.40	9.00	1.09
of Females	7.00	4.40	9.00	1.21
of Males	7.02	5.60	8.80	0.95
Attractiveness (Survey Measure)	0.01	-1.35	1.66	0.67
of Females	0.21	-1.35	1.66	0.72
of Males	-0.24	-1.19	0.81	0.51

M: Mean, *SD*: Standard Deviation

Table 4: Descriptive statistics

		Anonymous		Non-Anonymous	
Sex of Decision Maker	Sex of Co-Player	Predicted	Actual	Predicted	Actual
Female	Female			39.1	67.2
	Male			62.5	57.1
	Both	60.0	67.9	50.0	62.5
Male	Female			33.9	62.5
	Male			58.3	41.7
	Both	46.2	67.9	45.2	52.9
Both	Female			36.7	65.0
	Male			60.6	50.0
	Both	53.6	67.9	47.8	58.8

Table 5: Frequencies of predicted and actual market entries

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Predicted Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Financial Risk Attitude	0.017	-0.048	0.111	0.039	0.009	-0.104	0.025	0.282	-0.014	-0.054
	(0.042)	(0.052)	(0.067)	(0.072)	(0.087)	(0.082)	(0.128)	(0.160)	(0.123)	(0.070)
Anonymous Prediction (1 = *Entry*)	0.141	0.120	0.074	0.248	0.153	0.177	0.242	0.422	-0.221	-0.137
	(0.147)	(0.216)	(0.203)	(0.276)	(0.272)	(0.401)	(0.384)	(0.346)	(0.432)	(0.281)
Attractiveness	-0.035	-0.074	-0.019	0.137	-0.256	0.116	-0.342*	-0.039	0.008	0.137
	(0.068)	(0.094)	(0.137)	(0.106)	(0.136)	(0.135)	(0.174)	(0.188)	(0.261)	(0.129)
Attractiveness of Co-Player	0.109	0.186	0.043	0.212*	-0.125	0.430**	-0.338	0.043	0.100	0.267*
	(0.078)	(0.102)	(0.124)	(0.108)	(0.138)	(0.164)	(0.217)	(0.151)	(0.220)	(0.123)
Age	0.019	-0.001	0.024	-0.006	0.048	-0.054	0.088	-0.001	0.050	0.026
	(0.012)	(0.024)	(0.017)	(0.020)	(0.034)	(0.046)	(0.057)	(0.021)	(0.045)	(0.027)
Age of Co-Player	-0.001	0.005	-0.009	-0.029	-0.009	-0.019	0.023	-0.028	-0.032	-0.016
	(0.021)	(0.031)	(0.029)	(0.035)	(0.026)	(0.039)	(0.037)	(0.062)	(0.038)	(0.026)
Constant	-1.190	-0.815	-1.135	-2.296	1.840	-2.206	2.171	-1.033	-0.849	-2.899
	(1.046)	(1.513)	(1.625)	(1.556)	(1.840)	(2.348)	(2.729)	(2.319)	(2.874)	(1.627)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.018	0.025	0.018	0.056	0.054	0.143	0.140	0.063	0.032	0.052

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 6: Determinants of predictions (symmetry measure)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Prediction (1 = *Entry*)	-1.240***	-1.303***	-1.581***	-1.056***	-1.381***	-1.373**	-1.421**	-1.448**	-1.903***	-1.424***
	(0.224)	(0.280)	(0.378)	(0.300)	(0.329)	(0.421)	(0.485)	(0.459)	(0.525)	(0.312)
Financial Risk Attitude	0.018	-0.044	0.081	0.122	-0.113	0.059	-0.206	0.020	0.111	0.057
	(0.053)	(0.069)	(0.118)	(0.088)	(0.089)	(0.102)	(0.111)	(0.351)	(0.177)	(0.071)
Anonymous Decision (1 = *Entry*)	0.492*	0.422	0.572	0.177	0.834**	0.157	0.638	-0.295	1.583**	0.665
	(0.251)	(0.397)	(0.354)	(0.335)	(0.303)	(0.509)	(0.485)	(0.578)	(0.575)	(0.343)
Attractiveness	0.037	0.003	0.037	0.034	0.035	-0.066	0.069	0.071	-0.052	-0.008
	(0.130)	(0.184)	(0.227)	(0.152)	(0.143)	(0.214)	(0.170)	(0.230)	(0.338)	(0.169)
Attractiveness of Co-Player	-0.044	0.209	-0.348*	-0.084	-0.068	0.252	0.085	-0.498*	-0.516	0.101
	(0.093)	(0.128)	(0.153)	(0.107)	(0.158)	(0.148)	(0.266)	(0.239)	(0.264)	(0.120)
Age	0.026	-0.001	0.044	0.096*	-0.026	0.023	-0.029	0.265*	-0.038	-0.012
	(0.020)	(0.046)	(0.023)	(0.047)	(0.027)	(0.064)	(0.066)	(0.119)	(0.037)	(0.029)
Age of Co-Player	-0.004	0.018	-0.029	-0.031	0.010	-0.040	0.053	-0.038	-0.016	-0.039
	(0.021)	(0.030)	(0.026)	(0.032)	(0.031)	(0.039)	(0.049)	(0.058)	(0.047)	(0.026)
Confidence	0.087	0.074	0.110	0.126	0.049	0.151	-0.002	0.233	0.143	0.099
	(0.052)	(0.076)	(0.078)	(0.072)	(0.067)	(0.102)	(0.108)	(0.170)	(0.111)	(0.073)
Constant	-0.542	-1.417	1.318	-1.659	1.050	-1.016	-0.251	-2.868	3.930	0.244
	(1.362)	(2.571)	(1.885)	(1.776)	(1.806)	(2.696)	(3.116)	(2.993)	(2.820)	(1.809)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.173	0.187	0.245	0.173	0.230	0.203	0.234	0.327	0.328	0.233

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 7: Determinants of market entry decisions (symmetry measure)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Predicted Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Financial Risk Attitude	0.021	-0.047	0.114	0.064	-0.002	-0.072	0.038	0.282	-0.003	-0.025
	(0.042)	(0.051)	(0.062)	(0.071)	(0.087)	(0.086)	(0.133)	(0.159)	(0.115)	(0.071)
Anonymous Prediction (1 = *Entry*)	0.163	0.137	0.085	0.349	0.042	0.245	0.185	0.424	-0.168	-0.014
	(0.154)	(0.221)	(0.213)	(0.280)	(0.274)	(0.414)	(0.392)	(0.355)	(0.410)	(0.282)
Difference of Attractiveness Scores	-0.073	-0.126	-0.034	-0.054	-0.080	-0.142	-0.067	-0.041	-0.052	-0.071
	(0.049)	(0.073)	(0.073)	(0.075)	(0.090)	(0.104)	(0.128)	(0.111)	(0.151)	(0.086)
Age	0.018	-0.004	0.024	-0.014	0.049	-0.073	0.082	-0.001	0.050	0.020
	(0.012)	(0.024)	(0.017)	(0.020)	(0.034)	(0.043)	(0.059)	(0.021)	(0.045)	(0.026)
Age of Co-Player	-0.002	0.002	-0.009	-0.042	-0.009	-0.051	0.022	-0.028	-0.032	-0.021
	(0.021)	(0.030)	(0.029)	(0.036)	(0.026)	(0.036)	(0.042)	(0.062)	(0.039)	(0.027)
Constant	-0.644	0.094	-0.978	0.550	-0.750	2.749*	-2.477	-1.007	-0.161	0.035
	(0.631)	(0.944)	(0.827)	(1.063)	(1.102)	(1.369)	(2.108)	(1.824)	(1.300)	(0.859)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.010	0.021	0.018	0.025	0.027	0.060	0.052	0.063	0.031	0.012

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 8: Determinants of predictions and interaction of attractiveness scores (symmetry measure)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Prediction (1 = *Entry*)	-1.240***	-1.251***	-1.510***	-1.074***	-1.371***	-1.236**	-1.481**	-1.372**	-1.687***	-1.385***
	(0.224)	(0.302)	(0.389)	(0.302)	(0.328)	(0.443)	(0.460)	(0.467)	(0.506)	(0.311)
Financial Risk Attitude	0.017	-0.042	0.048	0.121	-0.113	0.068	-0.210	0.038	0.030	0.063
	(0.052)	(0.067)	(0.098)	(0.088)	(0.088)	(0.102)	(0.114)	(0.304)	(0.121)	(0.068)
Anonymous Decision (1 = *Entry*)	0.493*	0.348	0.484	0.184	0.830**	0.111	0.596	-0.278	1.156*	0.673*
	(0.251)	(0.405)	(0.351)	(0.338)	(0.306)	(0.504)	(0.496)	(0.554)	(0.550)	(0.343)
Difference of Attractiveness Scores	0.040	-0.088	0.215	0.059	0.050	-0.142	0.006	0.327*	0.203	-0.050
	(0.082)	(0.130)	(0.119)	(0.097)	(0.097)	(0.151)	(0.134)	(0.153)	(0.182)	(0.110)
Age	0.026	-0.004	0.041	0.097*	-0.027	0.015	-0.023	0.236*	-0.033	-0.014
	(0.020)	(0.045)	(0.024)	(0.048)	(0.027)	(0.064)	(0.066)	(0.107)	(0.038)	(0.029)
Age of Co-Player	-0.004	0.012	-0.026	-0.030	0.010	-0.049	0.051	-0.040	-0.015	-0.039
	(0.021)	(0.028)	(0.027)	(0.032)	(0.031)	(0.038)	(0.049)	(0.064)	(0.045)	(0.026)
Confidence	0.087	0.081	0.090	0.124	0.048	0.155	0.004	0.169	0.108	0.098
	(0.052)	(0.083)	(0.077)	(0.071)	(0.066)	(0.106)	(0.106)	(0.145)	(0.105)	(0.073)
Constant	-0.591	0.265	-0.590	-2.060	0.831	0.607	0.762	-4.847	0.509	0.908
	(0.843)	(1.698)	(0.932)	(1.467)	(1.151)	(1.827)	(2.541)	(2.908)	(1.725)	(1.097)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.173	0.177	0.231	0.173	0.230	0.195	0.230	0.304	0.300	0.232

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 9: Determinants of market entry decisions and interaction of attractiveness scores (symmetry measure)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Predicted Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Financial Risk Attitude	0.015	-0.049	0.117	0.052	-0.008	-0.112	0.072	0.296	-0.081	-0.045
	(0.042)	(0.062)	(0.071)	(0.075)	(0.087)	(0.105)	(0.137)	(0.162)	(0.125)	(0.072)
Anonymous Prediction (1 = *Entry*)	0.111	0.077	0.057	0.302	-0.006	0.104	0.058	0.355	-0.275	-0.080
	(0.157)	(0.230)	(0.229)	(0.275)	(0.270)	(0.380)	(0.384)	(0.373)	(0.395)	(0.275)
Attractiveness	0.006	0.017	0.067	0.085	-0.115	0.140	-0.191	0.254	-0.425	-0.123
	(0.173)	(0.212)	(0.244)	(0.236)	(0.233)	(0.325)	(0.326)	(0.396)	(0.421)	(0.241)
Attractiveness of Co-Player	-0.016	0.140	-0.169	0.183	-0.008	0.604**	-0.466	-0.143	0.610	0.365*
	(0.134)	(0.147)	(0.232)	(0.183)	(0.352)	(0.230)	(0.545)	(0.324)	(0.514)	(0.186)
Age	0.018	0.010	0.024	-0.007	0.042	-0.054	0.085	0.009	0.031	0.016
	(0.018)	(0.031)	(0.021)	(0.027)	(0.039)	(0.052)	(0.063)	(0.023)	(0.042)	(0.030)
Age of Co-Player	-0.002	0.017	-0.021	-0.033	-0.005	-0.012	-0.003	-0.046	0.001	0.004
	(0.025)	(0.033)	(0.038)	(0.041)	(0.034)	(0.038)	(0.051)	(0.070)	(0.051)	(0.030)
Constant	-0.591	-0.593	-0.674	0.173	-0.617	1.382	-2.016	-0.770	-0.120	-0.383
	(0.771)	(1.127)	(1.191)	(1.229)	(1.310)	(1.556)	(2.036)	(1.961)	(1.606)	(1.013)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.005	0.007	0.022	0.030	0.023	0.109	0.068	0.071	0.056	0.026

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 10: Determinants of predictions (survey measure)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Prediction (1 = *Entry*)	-1.260***	-1.244***	-1.626***	-1.069***	-1.407***	-1.154*	-1.393**	-1.622***	-1.620**	-1.401***
	(0.229)	(0.303)	(0.380)	(0.306)	(0.343)	(0.452)	(0.474)	(0.475)	(0.536)	(0.335)
Financial Risk Attitude	0.022	-0.070	0.039	0.131	-0.117	0.072	-0.256**	-0.136	0.097	0.058
	(0.051)	(0.073)	(0.092)	(0.090)	(0.086)	(0.112)	(0.097)	(0.263)	(0.158)	(0.080)
Anonymous Decision (1 = *Entry*)	0.481	0.406	0.738*	0.198	0.824**	0.205	0.810	0.298	1.065*	0.641
	(0.256)	(0.413)	(0.360)	(0.342)	(0.313)	(0.514)	(0.488)	(0.594)	(0.502)	(0.349)
Attractiveness	0.195	0.365	-0.567	0.099	0.327	0.399	0.360	-1.484**	0.388	0.477
	(0.189)	(0.233)	(0.351)	(0.289)	(0.234)	(0.374)	(0.296)	(0.496)	(0.672)	(0.302)
Attractiveness of Co-Player	-0.082	0.142	-0.321	-0.302	0.418	-0.091	0.654	-0.617**	0.013	0.023
	(0.140)	(0.206)	(0.195)	(0.160)	(0.345)	(0.254)	(0.498)	(0.224)	(0.635)	(0.230)
Age	0.041	0.030	0.011	0.103	-0.002	0.050	-0.015	0.137	-0.001	0.021
	(0.021)	(0.046)	(0.027)	(0.056)	(0.028)	(0.077)	(0.063)	(0.091)	(0.034)	(0.038)
Age of Co-Player	-0.008	0.035	-0.055	-0.047	0.038	-0.051	0.096	-0.017	-0.024	-0.027
	(0.024)	(0.031)	(0.032)	(0.036)	(0.037)	(0.048)	(0.060)	(0.060)	(0.055)	(0.033)
Confidence	0.085	0.086	0.135	0.120	0.049	0.149	0.018	0.282	0.073	0.093
	(0.053)	(0.079)	(0.080)	(0.075)	(0.068)	(0.106)	(0.114)	(0.150)	(0.104)	(0.073)
Constant	-0.844	-1.190	0.473	-1.764	-0.354	-0.340	-0.642	-3.291	0.015	-0.174
	(0.872)	(1.595)	(1.156)	(1.596)	(1.370)	(2.191)	(2.604)	(2.348)	(2.043)	(1.285)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.177	0.193	0.225	0.185	0.251	0.199	0.270	0.363	0.284	0.256

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 11: Determinants of market entry decisions (survey measure)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Predicted Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Financial Risk Attitude	0.015	-0.037	0.126	0.062	-0.010	-0.048	0.044	0.288	-0.091	-0.036
	(0.043)	(0.054)	(0.064)	(0.070)	(0.086)	(0.080)	(0.135)	(0.149)	(0.124)	(0.069)
Anonymous Prediction (1 = *Entry*)	0.111	0.088	0.061	0.323	0.003	0.214	0.153	0.356	-0.286	-0.055
	(0.157)	(0.228)	(0.225)	(0.275)	(0.268)	(0.396)	(0.392)	(0.374)	(0.392)	(0.274)
Difference of Attractiveness Scores	0.011	-0.052	0.140	-0.066	-0.078	-0.234	-0.018	0.171	-0.528	-0.243
	(0.116)	(0.140)	(0.161)	(0.151)	(0.212)	(0.205)	(0.285)	(0.259)	(0.399)	(0.170)
Age	0.018	0.005	0.029	-0.020	0.044	-0.080	0.092	0.004	0.025	0.005
	(0.014)	(0.026)	(0.020)	(0.021)	(0.038)	(0.046)	(0.062)	(0.021)	(0.039)	(0.027)
Age of Co-Player	-0.002	0.009	-0.018	-0.043	0.000	-0.046	0.028	-0.047	-0.005	-0.006
	(0.023)	(0.031)	(0.034)	(0.039)	(0.028)	(0.037)	(0.042)	(0.069)	(0.043)	(0.028)
Constant	-0.610	-0.309	-0.861	0.705	-0.800	2.731*	-2.876	-0.592	0.181	0.108
	(0.618)	(0.939)	(0.833)	(1.033)	(1.168)	(1.339)	(2.075)	(1.668)	(1.217)	(0.826)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.005	0.005	0.021	0.023	0.023	0.056	0.047	0.070	0.055	0.021

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 12: Determinants of predictions and interaction of attractiveness scores (survey measure)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: *Market Entry Decision* (1 = *Entry*)	All	Female	Male	Female Co-Player	Male Co-Player	Female vs. Female	Female vs. Male	Male vs. Female	Male vs. Male	Same Sex
Prediction (1 = *Entry*)	-1.259***	-1.189***	-1.489***	-1.087***	-1.382***	-1.065*	-1.483**	-1.375**	-1.647**	-1.322***
	(0.229)	(0.305)	(0.373)	(0.309)	(0.336)	(0.416)	(0.468)	(0.450)	(0.541)	(0.314)
Financial Risk Attitude	0.024	-0.041	0.119	0.126	-0.106	0.096	-0.218*	0.178	0.075	0.075
	(0.050)	(0.064)	(0.097)	(0.090)	(0.086)	(0.103)	(0.108)	(0.271)	(0.155)	(0.075)
Anonymous Decision (1 = *Entry*)	0.488	0.339	0.567	0.181	0.827**	0.197	0.590	0.020	1.136*	0.696*
	(0.252)	(0.420)	(0.363)	(0.337)	(0.316)	(0.512)	(0.501)	(0.558)	(0.536)	(0.344)
Difference of Attractiveness Scores	0.137	0.141	0.105	0.217	0.085	0.244	0.082	0.194	0.142	0.232
	(0.112)	(0.152)	(0.175)	(0.170)	(0.199)	(0.237)	(0.221)	(0.191)	(0.560)	(0.200)
Age	0.035	0.014	0.053*	0.114*	-0.020	0.039	-0.018	0.241*	-0.016	-0.003
	(0.020)	(0.043)	(0.022)	(0.053)	(0.026)	(0.070)	(0.065)	(0.105)	(0.034)	(0.032)
Age of Co-player	-0.013	0.009	-0.032	-0.040	0.002	-0.068	0.047	-0.013	-0.038	-0.050
	(0.022)	(0.031)	(0.028)	(0.038)	(0.030)	(0.049)	(0.045)	(0.061)	(0.045)	(0.031)
Confidence	0.084	0.077	0.087	0.124	0.044	0.139	0.001	0.159	0.092	0.087
	(0.053)	(0.080)	(0.077)	(0.073)	(0.065)	(0.105)	(0.104)	(0.141)	(0.104)	(0.074)
Constant	-0.602	-0.130	-1.031	-2.210	0.865	0.315	0.750	-6.221*	0.600	0.879
	(0.821)	(1.523)	(0.943)	(1.546)	(1.146)	(1.936)	(2.263)	(2.698)	(1.759)	(1.099)
Number of Observations	222	118	104	118	104	62	56	56	48	110
Number of Subjects	56	30	26	56	56	30	30	26	26	56
Pseudo R-squared	0.176	0.175	0.200	0.182	0.229	0.192	0.231	0.244	0.280	0.239

Probit regression with robust standard errors in parentheses, clustered at the individual level
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 13: Determinants of market entry decisions and interaction of attractiveness scores (survey measure)

The Higher Your Expectations, the Lower Your Trust: Avoiding the Experience of Unfulfilled Expectations

Wolfgang Breuer* Christiane Helduser* Philipp Schade*

Abstract

We study a binary-choice trust game and observe a negative relationship between people's expectations of others' trustworthiness and their own willingness to trust. This counterintuitive finding may be interpreted as the result of an avoidance strategy: The more people expect others to be trustworthy, the warier they are of their own expectations not being fulfilled, and the less willing they are to trust. Our research provides new insights into the origin of betrayal aversion, i.e. why people are less willing to take a risk if the source of the risk is a human being rather than a lottery. When facing human trustees, trustors tend to form expectations about their counterparts' behavior, and fit their own behavior to these expectations by lowering their willingness to trust. Presumably, this effect does not persist if trustees are replaced by non-human beings, i.e. lotteries, because the trustors' decision problem becomes more ambiguous in this case.

JEL Classification: A13, C91, D81

Keywords: Betrayal Aversion, Expectations, Trust, Trustworthiness

* Wolfgang Breuer, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 93539, Fax: +49 241 80 92163, E-Mail: wolfgang.breuer@bfw.rwth-aachen.de.

* Christiane Helduser, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 93561, Fax: +49 241 80 92163, E-Mail: christiane.helduser@bfw.rwth-aachen.de.

* Philipp Schade, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 94774, Fax: +49 241 80 92163, E-Mail: philipp.schade@bfw.rwth-aachen.de (corresponding author).

1 Introduction

It is widely undisputed that the experience of unfilled expectations generates negative emotions, such as disappointment or regret, because people tend to compare the factual outcome of a decision with counterfactual results which might have occurred if one had chosen differently (see Zeelenberg, van Dijk, Manstead, & van der Pligt, 2000 for an extensive overview). The anticipation of unfilled expectations is a powerful driver for human decision making, since people usually attempt to avoid the experience of negative emotions (van Dijk, Zeelenberg, & van der Pligt, 2003). In fact, there are several strategies for coping with the anticipated disconfirmation of expectations (Zeelenberg et al., 2000). First of all, people may intentionally underestimate the likelihood of obtaining a desired outcome in order to reduce negative emotions which they would experience if the desired outcome were not obtained (Cantor & Norem, 1989; Loewenstein & Linville, 1986; Pyszczynski, 1982; Shepperd, Ouellette, & Fernandez, 1996). Another simple but effective strategy is to avoid the risk of expectations being unfulfilled. In fact, choosing the safe option cannot even cause disappointment or regret, as the corresponding outcome does not vary and is known in advance. The tendency to avoid the experience of unfulfilled expectations can also be analyzed in the framework of prospect theory (Kahneman & Tversky, 1979) if decision makers are assumed to use their expectations as a reference point (Kőszegi & Rabin, 2006). More precisely, individuals may evaluate the outcome of a decision either as a gain or as a loss, depending on whether the outcome is in line with their initial expectations.

In the present paper, we investigate experimentally whether the anticipated experience of unfulfilled expectations operates in the context of trust, i.e. whether trustors lower their

willingness to trust, either because they attempt to avoid the risk of being betrayed or because they strategically overestimate the probability of being betrayed. As we will explain in the concluding section, the anticipated experience of unfulfilled expectations may be a source of betrayal aversion. However, our research goal is difficult to reach, because trustworthiness expectations and trust are related through different channels which may overlap and even counteract each other. We suggest that two key aspects have to be taken into account.

The first one is the empirical finding that more optimistic trustworthiness expectations increase trustors' expected utility of trusting and thus ceteris paribus their willingness to trust (Buchan, Croson, & Solnick, 2008). Previous studies which report a positive relationship between trustworthiness expectations and trust suggest indirectly that this effect outperforms the effect of anticipating unfulfilled expectations. Thus, identifying the effect of avoiding the experience of unfulfilled expectations requires a measure of trust that does not depend directly on trustworthiness expectations. Therefore, we employ the so-called *minimum acceptable probability* (*MAP*) trust game developed by Bohnet and Zeckhauser (2004). In Section 2, we will point out in more depth why this framework is adequate to study our research question. The second key aspect is the relationship between social preferences and trustworthiness expectations. Several authors find that trustors are prone to a social projection bias, i.e. social people expect higher returns when taking the trustor role than selfish people do (e.g. Bellemare & Kröger, 2007; Butler, Giuliano, & Guiso, 2015; Thielmann & Hilbig, 2014), because social people would themselves reward trust if they were trustees. The strong correlation between social preferences and trustworthiness expectations makes it difficult to separate the effect of anticipating unfulfilled expectations, since social preferences and the anticipation of unfulfilled

expectations may exert counteracting effects on trust. Social preferences increase one's willingness to trust, whereas we expect the contrary to be true for trustworthiness expectations, according to people's attempt to avoid the experience of unfulfilled expectations. Thus, measuring the effect of anticipating unfulfilled expectations requires disentangling it from the counteracting consequences of social preferences.

The remainder of this paper is organized as follows. In Section 2, we introduce the *MAP* trust game by Bohnet and Zeckhauser (2004) as well as the concept of social value orientation, which we use to operationalize social preferences. Furthermore, we present our experimental design and the procedures. In Section 3, we develop our hypotheses. This is followed by Section 4, where we describe and explain the findings obtained from the experiment. Section 5 summarizes our main results and points out some major theoretical implications.

2 Methods

2.1 Minimum Acceptable Probability Design

As outlined above, we build on previous work of Bohnet and Zeckhauser (2004), who study a variation of a binary-choice trust game (Camerer & Weigelt, 1988; Kreps, 1990). As depicted in Figure 1, the game is limited to three possible outcomes. If the trustor is not willing to trust ("no trust" alternative), both the trustor and the trustee receive a sure payoff, which we assume to be 10 points for each of them. If the trustor is willing to trust ("trust" alternative), her or his payoff is determined by the trustee. The trustee has to decide whether to reward ("reward" option) or to betray ("betrayal" option) the trustor. In the case of reward, the trustor and the trustee each receive the same payoff (15 points), which is higher than the first outcome. In the

case of betrayal, the trustor's payoff (8 points) is smaller than the sure payoff in the case of not trusting. The trustee's payoff is higher (22 points) in comparison to both the case of not trusting and the case of rewarding.

>>> Insert Figure 1 about here <<<

In the particular framework of Bohnet and Zeckhauser (2004), trustors have to state the probability at which they are indifferent between trusting and receiving the sure payoff, i.e. their *minimum acceptable probability* (*MAP*). The *MAP* represents trustors' switching point beyond which they are willing to take a risk by trusting rather than receiving the sure outcome by not trusting. An *MAP* below or equal to the proportion of trustees who reward trust, i.e. the actual reward probability (rp^a), indicates that trustors are willing to trust ($MAP \leq rp^a$), whereas an *MAP* above that proportion indicates the contrary ($MAP > rp^a$). Consequently, the higher a trustor's *MAP*, the less she or he is willing to trust. Trustors are informed about the actual reward probability only after they have reported their *MAP*.

The *MAP* design offers the important advantage that trustors' willingness to trust, measured in terms of their *MAPs*, should not be directly affected by their trustworthiness expectations. To explain this crucial aspect in a more rigorous way, we take a closer look at the trustors' decision problem. Assume $U(no\ trust)$ to be the utility derived from choosing the safe outcome and let $U(reward) > U(no\ trust)$ as well as $U(betrayal) < U(no\ trust)$ denote the utility if a trustor's trust is rewarded or betrayed, respectively. Thus, rational trustors have to balance the sure utility derived from not trusting ($U(no\ trust)$) and the expected utility from trusting by choosing their *MAPs* according to the following equation:

$$U(no\ trust) = MAP \cdot U(reward) + (1 - MAP) \cdot U(betrayal). \quad (1)$$

Setting an MAP^{+} below MAP according to (1), one would accept reward probabilities which lead to expected utilities below the utility of the sure outcome ($U(no\ trust)$). The contrary is true for $MAP^{+} > MAP$. In this case, one would forgo reward probabilities that generate beneficial expected utilities compared to $U(no\ trust)$. Since a trustor's MAP represents her or his minimum acceptable reward probability, the MAP should be independent of any considerations about the actual reward probability, i.e. her or his expected reward probability rp^{e}. However, as we outline in more depth in Section 3, this assumption has to be relaxed if people attempt to avoid the experience of unfulfilled expectations.

2.2 Social Value Orientation Slider Measure

As pointed out at the beginning, the effect of avoiding unfulfilled expectations can only be measured if trustors' concerns for trustees' payoffs, i.e. trustors' social preferences, are taken into account. We operationalize social preferences through the concept of social value orientation (SVO), which has been developed in the domain of social psychology in order to disentangle the intrinsic motivations of human behavior in situations where individuals have to decide how to allocate a scarce resource between themselves and others (Griesinger & Livingston, 1973; Liebrand, 1984). The SVO construct has been shown to be a valuable predictor of human behavior in social dilemmas (Balliet, Parks, & Joireman, 2009). In our experiment, trustors' *SVOs* are elicited by means of the *SVO Slider Measure* (Murphy, Ackermann, & Handgraaf, 2011), which provides a ranking order of a person's SVO, covering altruistic, competitive, proself, and prosocial *SVOs*. The *SVO Slider Measure* contains six so-called primary items. Each of them refers to a choice over the allocation of joint payoffs. To be more precise, each

item provides decision makers with a set of possible distributions and subjects have to indicate their preferred distribution. The mean allocations for self and for others derived from the primary items can be used to compute a single score of *SVO*, i.e. the so-called *SVO* angle. From an individual perspective, a higher *SVO* angle refers to a higher tendency toward perfect altruism. In comparison to earlier approaches, the *SVO Slider Measure* offers several advantages. First and foremost, it is a continuous measure and provides thus more accurate information on *SVO* than categorical approaches, such as the *SVO Ring Measure* (Liebrand, 1984). Second, it allows studying people's motivations in isolation, i.e. separating people's beliefs about others' behavior from their strategic concerns, as it is rooted in the decomposed game technique developed by Messick and McClintock (1968). This aspect is important for our purpose, because we intend to disentangle trustworthiness expectations and social preferences.

2.3 Experimental Design and Procedures

The experiment was conducted in the laboratory of a large German university, using z-Tree (Fischbacher, 2007). Subjects were recruited via ORSEE (Greiner, 2004). We conducted seven sessions with 26 subjects per session, obtaining 182 independent observations. Upon arrival, participants were randomly assigned to a computer. Once all participants were seated, they were informed that they could earn cash depending on their decisions and that they would be paid anonymously at the end of the experiment. The experiment itself consisted of one single treatment with three independent parts. First, subjects took part in the binary-choice trust game described in Section 2.1. Second and third, subjects' risk preferences and *SVOs* were elicited. Note that all three parts of the experiment were monetarily incentivized.

Given the high complexity of the binary-choice trust game, the experiment was designed with a strong focus on subjects' comprehension. At the beginning of the trust game part, subjects were provided with written instructions which they might use during the whole experiment. The experimental supervisor, being the same person in all sessions, read the instructions aloud and answered subjects' questions individually. Moreover, subjects were given plenty of time to study the instructions on their own. To ensure further that participants understood the *MAP* framework well, they had to pass a comprehension task in the style of Bohnet and Zeckhauser (2004). To be more precise, subjects had to calculate individual payoffs for trustors and trustees based on hypothetical *MAPs* and reward probabilities. Having answered all questions correctly, subjects were randomly assigned either to the trustor or to the trustee role. Roles were neutrally labeled. After the assignment of roles, trustors reported their *MAPs.* Simultaneously, trustees had to choose between the "reward" option and the "betrayal" option. Options were also neutrally labeled. The payoff structure in the three possible scenarios was the same as in Bohnet and Zeckhauser (2004) (see Section 2.1 for further details).

Subsequent to the *MAP* trust game, trustors faced the so-called *estimation task*, which had not been announced in the written instructions. In this task, trustors were asked to estimate the percentage of trustees that had chosen to reward in the binary-choice trust game, i.e. the expected reward probability or trustworthiness expectations. Subjects received an amount of two points (1.00 Euros) for a correct estimation. An estimation was defined to be correct if the deviation of the estimation from the actual percentage was equal to or smaller than five percentage points (upwards or downwards). If the estimation deviated by more than five percentage points and no more than ten percentage points, an amount of one point (0.50 Euros) was given to the subjects. We incentivized beliefs elicitation in this simple manner because the

procedure was easy to understand for subjects and did not require elaborated instructions. Subjects were not provided with any results from the trust game or any other information that might have helped them to solve the *estimation task* or to increase the expected payoff. Note that the *estimation task* was not part of the experiments conducted by Bohnet and Zeckhauser (2004). After the *estimation task*, each trustor was randomly paired with a trustee, and payoffs were calculated depending on the decisions made in the binary-choice trust game.

In the next part, subjects' risk aversion was elicited by means of the multiple price list method by Holt and Laury (2002). The multiple price list typically contains ten rows, each representing a choice between a "sure" and a "risky" gamble. After subjects had completed the multiple price list, one of their ten decisions was picked at random for remuneration. Subjects were informed about this procedure in advance. The result of the random draw was disclosed to participants only at the very end of the experiment. The multiple price list was followed by the *SVO Slider Measure* (Murphy et al., 2011), which has been outlined in more detail in Section 2.2. In a similar way to the remuneration of the multiple price list, only one of the decisions being part of the *SVO Slider Measure* was actually remunerated.

3 Hypotheses

The current section aims at providing a better understanding for the interplay between *SVO*, trustworthiness expectations, and the *MAP*. To identify the effect of avoiding the experience of unfulfilled expectations, it is crucial to consider the two channels through which an individual's *SVO* may affect her or his *MAP*. Figure 2 illustrates how both of these channels may operate in the binary-choice trust game that we study in our experiment. On the one hand,

SVO may have a direct impact on *MAP* (Path 1). On the other hand, the influence of *SVO* on *MAP* may be mediated by trustworthiness expectations (Paths 2 and 3), establishing an indirect relationship.

>>> Insert Figure 2 about here <<<

It is intuitive to assume that trustors' *SVOs* positively affect their willingness to trust, because trust inevitably increases trustees' payoffs, irrespective of trustees' choice, and also the overall welfare. This direct relationship between *SVO* and the *MAP* corresponds to Path 1. The indirect relationship, in contrast, is composed of two components (Paths 2 and 3). The first component (Path 2) links *SVO* to trustworthiness expectations and can be considered to be the result of social projection. Generally speaking, social projection is the tendency of individuals to extrapolate from their own behavior to the behavior of others (e.g. Robbins & Krueger, 2005). Numerous studies provide evidence for the relationship between one's trustworthiness expectations and one's willingness to trust being consistent with social projection (e.g. Bellemare & Kröger, 2007; Butler et al., 2015; Thielmann & Hilbig, 2014).

The second component (Path 3) links an individual's trustworthiness expectations to her or his *MAP*, and refers thus to the hypothesized effect of avoiding the experience of unfulfilled expectations. As outlined earlier, disconfirmed expectations constitute a source of disutility, because trustors use their trustworthiness expectations as a reference point when evaluating the possible outcomes of the trust game. In this aspect, we extent previous research by Bohnet, Herrmann, and Zeckhauser (2010) and Breuer, Helduser, and Schade (2016), who emphasize the importance of reference points in the *MAP* trust game.

More formally, we argue that for trustors who set their *MAPs* according to (1), $U(betrayal)$ is a decreasing function of their trustworthiness expectations rp^e. As a consequence, trustors with high subjective estimates rp^e, i.e. optimistic trustors, are inclined to increase their *MAPs* in order to circumvent disutility arising from unfulfilled expectations in the case of betrayal. To be more precise, according to (1) and to the implicit function theorem, we obtain the following relationship for $0 < MAP < 1$ and $U(reward) > U(betrayal)$:

$$\frac{\partial MAP}{\partial U(betrayal)} = -\frac{1 - MAP}{U(reward) - U(betrayal)} < 0 \tag{2}$$

From (2), we may conclude that $\partial MAP/\partial rp^e = \partial MAP/\partial U(betrayal) \cdot \partial U(betrayal)/\partial rp^e > 0$ due to $\partial U(betrayal)/\partial rp^e < 0$.

As outlined above, the two components of the indirect effect differ in their signs. Thus, the corresponding relationship between *SVO* and the *MAP* is positive. Since the direct and the indirect effect of *SVO* on the *MAP* counteract each other, the sign of the total effect, i.e. the aggregate of the direct and the indirect effect, is a priori unclear. However, the existing literature suggests a positive relationship between an individual's *SVO* and her or his willingness to trust (Kanagaretnam, Mestelman, Nainar, & Shehata, 2009), which refers to a negative total effect of *SVO* on the *MAP* in our context. For similar reasons, the total effect of trustworthiness expectations on the *MAP* is unclear. On the one hand, high expectations may induce low *MAPs* (Path 3). On the other hand, regardless of the causal direction, expectations are positively correlated with *SVOs*, due to social projection (Path 2). Therefore, there is also a positive indirect relationship between expectations and *MAP* (Paths 1 and 2). As the effect of avoiding the experience of unfulfilled expectations conflicts with the consequences for trust as a function

of a trustor's *SVO* (Path 1), the total empirical effect regarding the relationship between expectations and the *MAP* remains unclear, as long as Path 3 is not separated from Paths 1 and 2. From a more formal point of view, identifying the effect of avoiding the experience of unfulfilled expectations in a regression model requires that *SVO* is controlled for, and vice versa. Up to this point, our reasoning can be mapped to three hypotheses, each of which refers to a path in Figure 2:

Hypothesis 1: *There is a positive relationship between trustors' SVOs and their willingness to trust, especially if their trustworthiness expectations are controlled for (cf. Path 1).*

Hypothesis 2: *There is a positive relationship between trustors' SVOs and their trustworthiness expectations (cf. Path 2).*

Hypothesis 3: *There is a negative relationship between trustors' trustworthiness expectations and their willingness to trust, especially if their SVOs are controlled for (cf. Path 3).*

We have suggested that the attempt to avoid unfulfilled expectations constitutes a positive relationship between trustworthiness expectations and the *MAP*. However, it is natural to assume that this effect is driven by people expecting the "trust" alternative to be chosen, i.e. people indicating low *MAPs* or holding optimistic trustworthiness expectations. These people are more concerned with avoiding negative emotions, because the "trust" alternative bears a higher risk of not being satisfied with the outcome than the "no trust" alternative does. In the case of betrayal, people know for sure that the "no trust" alternative, which could have been reached by choosing higher *MAPs*, would have yielded a higher payoff. In contrast, people who

end up with the "no trust" alternative will not find out whether the "trust" alternative would have actually been a better choice or not. Compared to the "no trust" alternative, choosing the "trust" alternative produces worse outcomes for trustors if trustees decide to betray. This reasoning corresponds to previous research in a broader sense. Zeelenberg (1999), for instance, suggests that anticipated regret increases if people expect to get feedback on their choice. Thus, we propose the following hypothesis:

Hypothesis 4: *The negative relationship between people's trustworthiness expectations and their willingness to trust is more intense for trustors who expect the "trust" alternative to be chosen than it is for others.*

4 Results

4.1 Descriptive Statistics

Across all sessions, 91 subjects were assigned to the trustor role (54 males and 37 females) and to the trustee role (51 males and 40 females), respectively, and the proportion of female participants ranged from 30.8 % to 53.8 %. Trustors' mean *MAP* was 57.2 % ($SD = 0.259$). The *MAPs* did not differ significantly either between female and male trustors (Mann-Whitney test, $Z = -0.961$, $p = 0.336$) or between sessions (Chi-squared test, $\chi^2 = 131.723$, $p = 0.760$). The overall actual reward probability rp^a, computed across all sessions, was 49.5 %. The corresponding statistics are depicted in Table 1.

>>> Insert Table 1 about here <<<

According to the *estimation task*, trustors expected on average 47.8 % ($SD = 0.243$) of the trustees to be willing to reward trust. Thus, trustors only slightly underestimated the actual reward probability ($rp^a = 49.5$ %). Referring to their trustworthiness expectations, women ($M = 39.2$ %, $SD = 0.212$) were considerably more pessimistic than men were ($M = 53.7$ %, $SD = 0.247$). An independent-samples *t*-test ($t(89) = -2.911$, $p < 0.01$) shows that the difference of the trustworthiness expectations between female and male trustors was statistically significant. This finding is in line with previous research on gender differences in preferences (e.g. Croson & Gneezy, 2009; Garbarino & Slonim, 2009). The corresponding statistics are depicted in Table 2.

>>> Insert Table 2 about here <<<

The *SVO Slider Measure* revealed that trustors can be classified either as being proself (49.5 %) or prosocial (50.5 %). According to a Chi-squared test ($\chi^2 = 0.016$, $p = 0.899$), the proportion of prosocial subjects did not differ significantly between female trustors (48.6 %) and male trustors (50.0 %). A more rigorous way to describe a subject's *SVO* is to consider her or his continuous *SVO* angle, computed by the *SVO Slider Measure*. Trustors' average angle was 20.946 degrees ($SD = 13.087$). The mean *SVO* angle was lower for women ($M = 19.757$, $SD = 13.146$) than for men ($M = 21.761$, $SD = 13.107$). A Mann-Whitney test did not indicate a statistically significant gender difference in this aspect ($Z = -1.056$, $p = 0.294$). As the *SVO* angle allows values ranging from -16.26 to 61.39 degrees and the threshold between proself and prosocial *SVO* is 22.45 degrees, the mean *SVO* angles indicate a slight tendency toward proself *SVO*.

4.2 Hypotheses Testing

Our hypotheses suggest that the effect of SVO on trust is mediated by trustworthiness expectations. The effect of anticipating unfulfilled expectations (Path 3 in Figure 2) is part of the indirect effect which relates SVO to trust (Paths 2 + 3). We expect the indirect effect to mitigate the total effect of SVO (the so-called suppressor effect). An indirect effect is said to be a suppressor effect if the direct and the indirect effect that an independent variable exerts on a dependent variable have opposite signs. As we study a standard mediation setting containing an independent variable (SVO), a mediator variable (*expected reward probability/trustworthiness expectations* rp^e), and a dependent variable of trust (MAP), we apply the stepwise approach established by Baron and Kenny (1986) to verify our hypotheses. In addition, we explicitly test the statistical significance of the indirect effect. Since the variables of interest are fractions, we base the following analysis on generalized linear models with logistic link functions (Papke & Wooldridge, 1996).

In the first step, we consider the relationship between SVO and the MAP without controlling for the mediator variable, i.e. the total effect of SVO. In all models, we control for subjects' gender, because people's behavior may differ between women and men (e.g. Croson & Gneezy, 2009; Garbarino & Slonim, 2009). Model 1 (see Table 3) reveals a significantly negative effect: The more socially orientated a trustor is, the lower her or his MAP is. Note that a lower MAP implies a higher willingness to trust, and vice versa. This finding corresponds to Hypothesis 1. In the second step, we study the first component of the indirect effect of SVO on trust (Path 2). Figure 3 illustrates that, in line with Hypothesis 2, prosocial trustors have higher mean trustworthiness expectations rp^e ($M = 0.571,\ SD = 0.233$) than proself trustors

have ($M = 0.388$, $SD = 0.220$). An independent-samples t-test shows that trustworthiness expectations differ significantly between prosocial and proself trustors ($t(89) = -3.844$, $p < 0.001$). According to the assumption that subjects are prone to social projection, prosocial trustors overestimate the reward probability, whereas proself trustors underestimate it. Model 2 of Table 3 confirms the positive effect of SVO on *trustworthiness expectations*.

>>> Insert Table 3 and Figure 3 about here <<<

In the third step, we examine the second component of the indirect effect (Path 3). We observe a positive effect of trustworthiness expectations on the MAP (Model 3) which is significant on the 10 % level. To separate the effect of Path 3 from the counteracting effect of SVO on the MAP (Path 1), we run an additional model in which we control for SVO (Model 4). In the extended model, the effect of trustworthiness expectations becomes highly significant. The effect size of both trustworthiness expectations and SVO substantially increases. These findings are in line with our theoretical considerations, since the direct effect and the indirect effect are expected to mitigate each other. These results correspond to Hypotheses 1 and 3. Note that we obtain similar results by running linear regressions in the style of Bohnet and Zeckhauser (2004) (see Table A.1 in the Appendix).

To complete the mediation analysis, we test the statistical significance of the indirect effect. A Sobel test reveals that the mediation effect of trustworthiness expectations is statistically significant ($p = 0.020$), reducing the size of the total effect of SVO on trust by approximately 52.2 %. As the Sobel test is often criticized for its low power (see Zhao, Lynch, & Chen, 2010 for further details), we also test the indirect effect by means of the more rigorous bootstrapping method (Preacher & Hayes, 2004). Using 10,000 resamples, bootstrapping also

indicates a statistically significant indirect effect ($p = 0.019$). Summing up, both tests confirm the results obtained from our stepwise analysis.

In order to test Hypothesis 4, we split our dataset into two subsamples, based on whether subjects expected the "no trust" alternative ($MAP > rp^e$) or the "trust" alternative ($MAP \leq rp^e$) to be chosen. We separately rerun Model 4 for both subsamples to identify potential differences with respect to the role of trustworthiness expectations. According to Models 5 and 6, the effect of trustworthiness expectations on the MAP is significantly positive in both subsamples. However, the respective coefficients indicate that the effect is more pronounced if people expect the "trust" alternative to be chosen. In fact, the coefficients differ significantly between the two subsamples ($\chi^2 = 332.80$, $p < 0.001$). This finding, which is fully in line with Hypothesis 4, can be confirmed by the OLS regressions presented in the Appendix. The substantial increase of the adjusted R-squared from Model 5 to Model 6 (see Table A.1 in the Appendix) provides additional support for the elevated importance of trustworthiness expectations if the "trust" alternative is expected to be chosen.

Since trust is often assumed to be a risky decision (e.g. Eckel & Wilson, 2004), we repeat the previous steps and control for subjects' risk aversion, which we measure by means of the number of "safe" options chosen in the multiple price list (Holt & Laury, 2002). Note that all those subjects holding inconsistent risk preferences, including subjects switching more than once between the two options and those rejecting the safe option in the last decision problem, have to be dropped. The inconsistency rate of 18.7 % (17 out of 91 subjects) is in line with the literature (see, e.g., Holt & Laury, 2002). The extended models reassure our previous findings in all aspects (see Model 1 to Model 6 in Table 4), yielding also highly significant effects. Again, the coefficients of trustworthiness expectations differ significantly between both subsamples

($\chi^2 = 302.52$, $p < 0.001$). Note that the corresponding OLS regressions are presented in Table A.2 in the Appendix.

>>> Insert Table 4 about here <<<

In analogy to our previous analysis, we also re-conduct the Sobel test. If risk aversion is included, the mediation effect of trustworthiness expectations turns out to be statistically significant ($p = 0.028$). The indirect effect of trustworthiness expectations on trust reduces the size of the total effect by approximately 66.7 %. The more rigorous bootstrapping approach with 10,000 resamples confirms the statistical significance of the indirect effect ($p = 0.038$).

4.3 Causality

Our previous analysis revealed a positive correlation between subjects' trustworthiness expectations and their *MAPs*. However, we did not discuss the two distinct causalities that may underlie this relationship: People may either truthfully reveal their *MAPs* and fit their trustworthiness expectations to the latter, or the other way round. If trustors are biased in terms of their expectations, one would expect their average expectations to be lower than those of non-biased subjects, as a consequence of avoiding the experience of unfulfilled expectations. As an external validation of trustworthiness expectations, we compare trustors' expectations to those of two distinct reference groups which we expect to be unbiased in this regard. First, we compare trustors' expectations ($M = 47.8$ %, $SD = 0.243$, $N = 91$) with those reported by trustees ($M = 46.9$ %, $SD = 0.264$, $N = 91$). Second, we consider trustworthiness expectations of people who have not even participated in the experiment.

To obtain trustworthiness expectations from outside the laboratory, we conducted an online survey among students from a large German university. The survey was advertised in several undergraduate and postgraduate courses. Analogously to the laboratory experiment, participants ($N = 120$) were first provided with general information on the binary-choice trust game, such as the payoff matrix. Subsequently, they had to pass a comprehension test in the style of Bohnet and Zeckhauser (2004). Subjects could only proceed if they had correctly answered all comprehension questions. Then, they were asked to estimate the reward probability which we had actually observed in the laboratory experiment ($rp^a = 49.5\ \%$). The estimation was incentivized in a similar manner to the *estimation task* in our laboratory experiment. Participants received an additional fee of 5.00 Euros. The total amounts were transferred to participants in the form of a digital gift voucher for the university shop.

Comparing the three samples in terms of subjects' trustworthiness expectations confirms that trustors fit their *MAPs* to their expectations, and not the other way round. Since the data from the three samples are not normally distributed (see Table 5 for descriptive statistics), we carry out a set of non-parametrical tests. A Kruskal-Wallis test does not reveal a statistically significant difference between the three samples ($H(2) = 1.702,\ p = 0.427$). To account for potential gender differences, we split all three samples with respect to subjects' gender and run separate Kruskal-Wallis tests for women and men. The additional tests do not indicate significant differences either within the female subsample ($H(2) = 2.756,\ p = 0.252$) or the male subsample ($H(2) = 3.641,\ p = 0.162$).

>>> Insert Table 5 about here <<<

5 Conclusion

In this paper we have studied the relationship between trustors' expectations about others' trustworthiness and their own willingness to trust. In contrast to previous studies, we found that the more people expect others to be trustworthy, the *less* willing they are to trust. This counterintuitive finding can be interpreted as follows: People lower their willingness to trust in order to avoid the experience of negative emotions, such as disappointment or regret, which will occur if positive expectations are not fulfilled. As we could show experimentally, measuring the anticipation of unfulfilled expectations in the context of trust requires a particular experimental design. First, the measure for people's willingness to trust should not depend directly on their trustworthiness expectations. For this reason, we employed the *MAP* trust game by Bohnet and Zeckhauser (2004), which is based on the latent assumption of trust measured in terms of people's *MAPs* being independent of trustworthiness expectations (Bohnet et al., 2010; Bohnet, Greig, Herrmann, & Zeckhauser, 2008). Second, social preferences have to be taken into account, since they are entangled with both trust and trustworthiness expectations.

From a theoretical point of view, our findings are intriguing, since they provide new insights into the origin of betrayal aversion. Betrayal aversion refers to the phenomenon that people are less willing to take a risk if the source of the risk is a human being rather than a lottery (Aimone & Houser, 2012; Bohnet et al., 2008; Bohnet et al., 2010; Bohnet & Zeckhauser, 2004). A common approach to demonstrate the existence of betrayal aversion is to compare the binary-choice trust game of Bohnet and Zeckhauser (2004) to a so-called risky dictator game, where the human trustee is replaced by a synthetic trustee, i.e. a lottery. The existence of betrayal aversion is assumed for situations with trustors stating higher *MAPs* in the binary-

choice trust game than they do in the risky dictator game. A reason for betrayal aversion is that people's perceptions of structurally identical payoffs depend on whether these payoffs are determined by a human being or a lottery: A monetary loss in the binary-choice trust game is regarded as an act of betrayal, whereas such a loss in a lottery is merely evaluated as "bad luck" (Fetchenhauer & Dunning, 2009).

The present study provides empirical evidence for why people are betrayal averse. In line with Aimone and Houser (2012), who briefly point out to loss aversion as a potential driver for betrayal aversion, we argue that the observed effect of avoiding unfulfilled expectations is likely to occur in the binary-choice trust game rather than in the risky dictator game. In the binary-choice trust game, the probabilities of reward and betrayal are ex ante unknown. Therefore, the decision about whether to trust can be considered to be an ambiguous one. However, people tend to form beliefs about the probabilities of reward and betrayal, e.g. based on past experience or on social projection. In this aspect, the binary-choice trust game differs substantially from the risky dictator game, where neither past experience nor social projection provide valuable information on how the lottery will decide. Thus, in comparison to the binary-choice trust game, the risky dictator game resembles even more a decision problem under ambiguity, and is less prone to the effect of avoiding unfulfilled expectations. According to this line of reasoning, betrayal aversion is not a direct consequence of facing a human antagonist per se, but rather caused by the fact that human behavior might be anticipated in an easier way than random reactions by computers. Taken together, our findings may at least partly explain why people state higher *MAPs* in the binary-choice trust game than they do in the risky dictator game. Future research should address whether the different drivers of betrayal aversion can be disentangled from each other.

Acknowledgments

We thank Anselm Hüwe, Stefan Kleeschulte, Christine Stibbe, participants in the 2016 annual meeting of the German Association for Experimental Economic Research (GfeW) in Gießen, Germany, as well as participants in the MOE Doctoral Workshop at RWTH Aachen University for valuable comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aimone, J. A., & Houser, D. (2012). What You Don't Know Won't Hurt You: A Laboratory Analysis of Betrayal Aversion. *Experimental Economics*, *15*(4), 571–588.

Balliet, D., Parks, C., & Joireman, J. (2009). Social Value Orientation and Cooperation in Social Dilemmas: A Meta-Analysis. *Group Processes and Intergroup Relations*, *12*(4), 533–547.

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bellemare, C., & Kröger, S. (2007). On Representative Social Capital. *European Economic Review*, *51*(1), 183–202.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, *98*(1), 294–310.

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). Trust and the Reference Points for Trustworthiness in Gulf and Western Countries. *Quarterly Journal of Economics*, *125*(2), 811–828.

Bohnet, I., & Zeckhauser, R. (2004). Trust, Risk and Betrayal. *Journal of Economic Behavior and Organization*, *55*(4), 467–484.

Breuer, W., Helduser, C., & Schade, P. (2016). Breaking the Rules: Anticipation of Norm Violation in a Binary-Choice Trust Game. *Economics Letters*, *146*, 123–125.

Buchan, N. R., Croson, R. T. A., & Solnick, S. (2008). Trust and Gender: An Examination of Behavior and Beliefs in the Investment Game. *Journal of Economic Behavior and Organization*, *68*(3), 466–476.

Butler, J. V., Giuliano, P., & Guiso, L. (2015). Trust, Values, and False Consensus. *International Economic Review*, *56*(3), 889–915.

Camerer, C., & Weigelt, K. (1988). Experimental Tests of a Sequential Equilibrium Reputation Model. *Econometrica*, *56*(1), 1–36.

Cantor, N., & Norem, J. K. (1989). Defensive Pessimism and Stress and Coping. *Social Cognition*, *7*(2), 92–112.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Eckel, C. C., & Wilson, R. K. (2004). Is Trust a Risky Decision? *Journal of Economic Behavior and Organization*, *55*(4), 447–465.

Fetchenhauer, D., & Dunning, D. (2009). Do People Trust Too Much or Too Little? *Journal of Economic Psychology*, *30*(3), 263–276.

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

Garbarino, E., & Slonim, R. (2009). The Robustness of Trust and Reciprocity Across a Heterogeneous US Population. *Journal of Economic Behavior and Organization*, *69*(3), 226–240.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In V. Macho & Kremer K. (Eds.), *Forschung und wissenschaftliches Rechnen* (63rd ed., pp. 79–93). Göttingen: Gesellschaft für wissenschaftliche Datenverarbeitung.

Griesinger, D. W., & Livingston, J. W. (1973). Toward a Model of Interpersonal Motivation in Experimental Games. *Systems Research and Behavioral Science*, *18*(3), 173–188.

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, *92*(5), 1644–1655.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, *47*(2), 263–291.

Kanagaretnam, K., Mestelman, S., Nainar, K., & Shehata, M. (2009). The Impact of Social Value Orientation and Risk Attitudes on Trust and Reciprocity. *Journal of Economic Psychology*, *30*(3), 368–380.

Kőszegi, B., & Rabin, M. (2006). A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics*, *121*(4), 1133–1165.

Kreps, D. M. (1990). Corporate Culture and Economic Theory. In J. E. Alt & K. A. Shepsle (Eds.), *Perspectives on Positive Political Economy* (pp. 90–143). Cambridge: Cambridge University Press.

Liebrand, W. B. G. (1984). The Effect of Social Motives, Communication and Group Size on Behaviour in an N-Person Multi-Stage Mixed-Motive Game. *European Journal of Social Psychology*, *14*(3), 239–264.

Loewenstein, G. & Linville, P. (1986). Expectation Formation and the Timing of Outcomes: Cognitive Strategy for Balancing the Conflicting Incentives for Savoring Success and Avoiding Disappointment (*Unpublished Manuscript*).

Messick, D. M., & McClintock, C. G. (1968). Motivational Bases of Choice in Experimental Games. *Journal of Experimental Social Psychology*, *4*(1), 1–25.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring Social Value Orientation. *Judgment and Decision Making*, *6*(8), 771–781.

Papke, L. E., & Wooldridge, J. M. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics*, *11*(6), 619–632.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models. *Behavior Research Methods, Instruments, and Computers*, *36*(4), 717–731.

Pyszczynski, T. (1982). Cognitive Strategies for Coping with Uncertain Outcomes. *Journal of Research in Personality*, *16*(3), 386–399.

Robbins, J. M., & Krueger, J. I. (2005). Social Projection to Ingroups and Outgroups: A Review and Meta-Analysis. *Personality and Social Psychology Review*, *9*(1), 32–47.

Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning Unrealistic Optimism: Performance Estimates and the Temporal Proximity of Self-Relevant Feedback. *Journal of Personality and Social Psychology*, *70*(4), 844.

Thielmann, I., & Hilbig, B. E. (2014). Trust in Me, Trust in You: A Social Projection Account of the Link Between Personality, Cooperativeness, and Trustworthiness Expectations. *Journal of Research in Personality*, *50*, 61–65.

van Dijk, W. W., Zeelenberg, M., & van der Pligt, J. (2003). Blessed are Those Who Expect Nothing: Lowering Expectations as a Way of Avoiding Disappointment. *Journal of Economic Psychology*, *24*(4), 505–516.

Zeelenberg, M. (1999). Anticipated Regret, Expected Feedback and Behavioral Decision Making. *Journal of Behavioral Decision Making*, *12*(2), 93–106.

Zeelenberg, M., van Dijk, W. W., Manstead, A. S. R., & van der Pligt, J. (2000). On Bad Decisions and Disconfirmed Expectancies: The Psychology of Regret and Disappointment. *Cognition and Emotion*, *14*(4), 521–541.

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths About Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206.

Tables

	Minimum Acceptable Probabilities (*MAPs*)		
	Overall	Women	Men
Mean	0.572	0.598	0.553
Median	0.600	0.600	0.600
Standard Deviation	0.259	0.249	0.267
N	91	37	54

Table 1: Minimum acceptable probabilities stated in the binary-choice trust game

	Trustworthiness Expectations (rp^e)		
	Overall	Women	Men
Mean	0.478	0.392	0.537
Median	0.470	0.400	0.575
Standard Deviation	0.243	0.212	0.247
N	91	37	54

Table 2: Trustors' trustworthiness expectations stated in the estimation task

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	MAP	rp^e	MAP	MAP	MAP ($MAP > rp^e$)	MAP ($MAP \leq rp^e$)
Social Value Orientation (SVO)	-0.020***	0.031***		-0.031***	-0.003	-0.022**
	(0.007)	(0.005)		(0.008)	(0.010)	(0.009)
Expected Reward Probability (rp^e)			0.682*	1.446***	1.764***	4.420***
			(0.398)	(0.430)	(0.292)	(1.093)
Gender (0 = *female*, 1 = *male*)	-0.146	0.549***	-0.283	-0.333*	-0.061	-0.540***
	(0.187)	(0.134)	(0.203)	(0.183)	(0.237)	(0.137)
Constant	0.798***	-1.072***	0.132	0.466*	0.215	-2.203***
	(0.193)	(0.132)	(0.212)	(0.243)	(0.211)	(0.840)
Observations	91	91	91	91	54	37

Generalized Linear Models with logit link functions
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: Influences on the dependent variables *MAP* and *Expected Reward Probability*

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	MAP	rp^e	MAP	MAP	MAP ($MAP > rp^e$)	MAP ($MAP \leq rp^e$)
Social Value Orientation (SVO)	-0.016***	0.032***		-0.028***	-0.004	-0.011**
	(0.005)	(0.010)		(0.005)	(0.012)	(0.005)
Expected Reward Probability (rp^e)			0.747	1.502***	2.097***	4.418***
			(0.499)	(0.475)	(0.800)	(1.182)
Gender (0 = *female*, 1 = *male*)	0.004	0.668***	-0.200	-0.221	-0.051	-0.429***
	(0.248)	(0.097)	(0.251)	(0.228)	(0.183)	(0.163)
Risk Aversion	0.123*	0.036	0.130	0.112	0.169***	0.012
	(0.070)	(0.033)	(0.082)	(0.069)	(0.034)	(0.099)
Constant	0.071	-1.271***	-0.553	-0.230	-0.840***	-2.443***
	(0.500)	(0.291)	(0.619)	(0.516)	(0.313)	(0.639)
Observations	74	74	74	74	45	29

Generalized Linear Models with logit link functions
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Influences on the dependent variables *MAP* and *Expected Reward Probability*

		Trustworthiness Expectations		
Source	Sample	Overall	Women	Men
Experiment	Trustors	0.478	0.392	0.537
Experiment	Trustees	0.461	0.491	0.452
Online Survey	All Participants	0.437	0.411	0.456

Table 5: Mean trustworthiness expectations

Figures

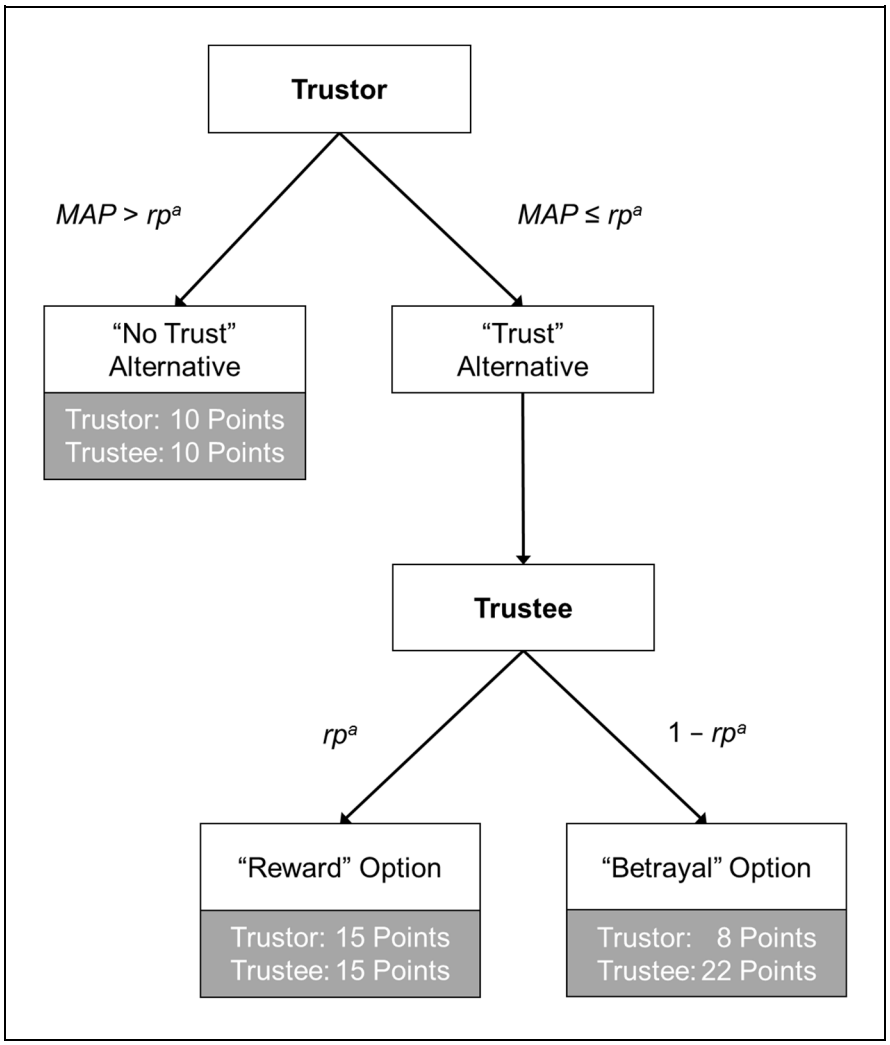

Figure 1: Binary-choice trust game

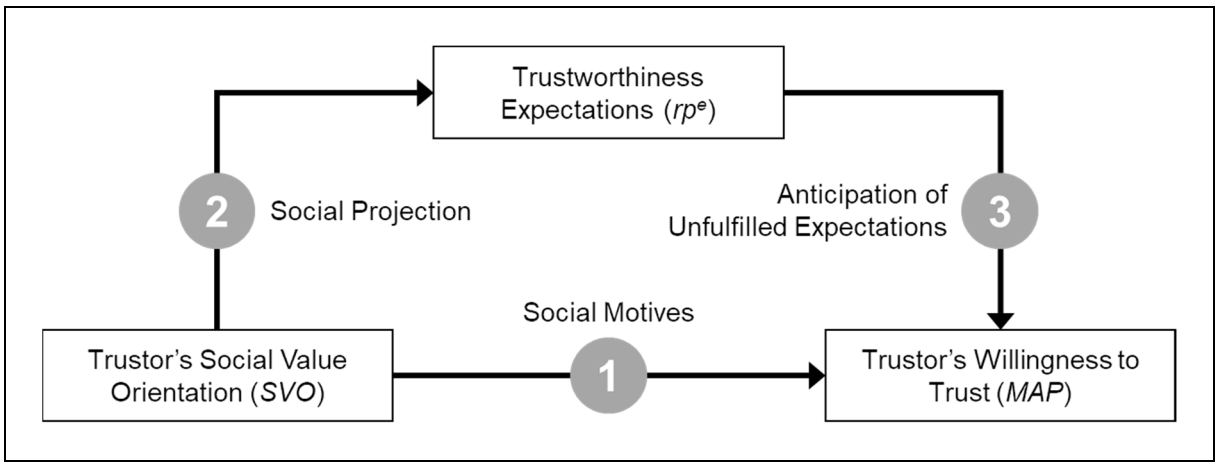

Figure 2: Direct and indirect effect of SVO on the MAP

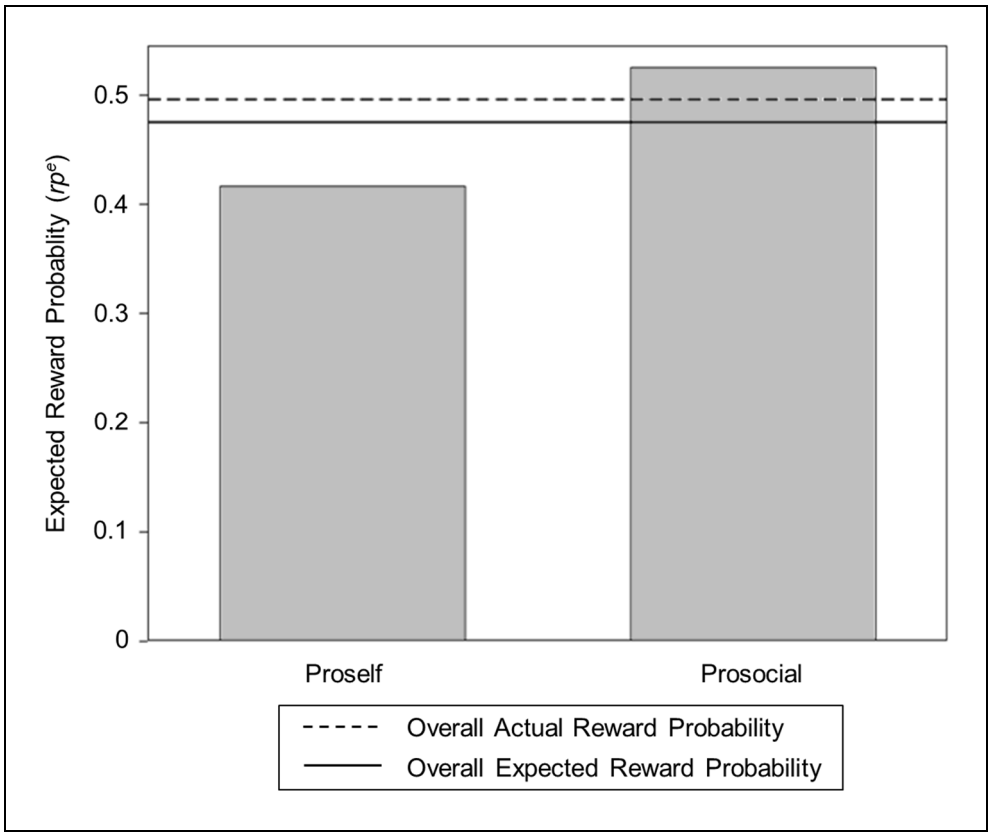

Figure 3: Actual and expected reward probabilities on the overall level

Appendix

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	MAP	rp^e	MAP	MAP	MAP ($MAP > rp^e$)	MAP ($MAP \leq rp^e$)
Social Value Orientation (SVO)	-0.005**	0.007***		-0.007***	-0.001	-0.004
	(0.002)	(0.001)		(0.002)	(0.002)	(0.003)
Expected Reward Probability (rp^e)			0.166	0.339**	0.363***	0.802***
			(0.097)	(0.100)	(0.070)	(0.159)
Gender (0 = *female*, 1 = *male*)	-0.035	0.130***	-0.069	-0.079	-0.012	-0.072*
	(0.045)	(0.033)	(0.050)	(0.044)	(0.050)	(0.032)
Constant	0.693***	0.245***	0.533***	0.610***	0.564***	0.064
	(0.044)	(0.030)	(0.052)	(0.057)	(0.050)	(0.183)
Observations	91	91	91	91	54	37
Adj. R-squared	0.045	0.229	0.008	0.113	0.091	0.407

OLS regressions
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.1: Influences on the dependent variables *MAP* and *Expected Reward Probability* (OLS regressions)

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	MAP	rp^e	MAP	MAP	MAP ($MAP > rp^e$)	MAP ($MAP \leq rp^e$)
Social Value Orientation (SVO)	-0.004**	0.007**		-0.006***	-0.000	-0.002
	(0.001)	(0.002)		(0.001)	(0.002)	(0.001)
Expected Reward Probability (rp^e)			0.173	0.339**	0.396*	0.985***
			(0.117)	(0.106)	(0.169)	(0.254)
Gender (0 = *female*, 1 = *male*)	0.001	0.159***	-0.047	-0.053	-0.012	-0.087*
	(0.059)	(0.022)	(0.060)	(0.055)	(0.038)	(0.041)
Risk Aversion	0.029	0.008	0.031	0.026	0.034***	0.006
	(0.017)	(0.008)	(0.020)	(0.016)	(0.008)	(0.024)
Constant	0.519***	0.207**	0.373**	0.449**	0.358***	-0.068
	(0.123)	(0.066)	(0.150)	(0.122)	(0.074)	(0.149)
Observations	74	74	74	74	45	29
Adj. R-squared	0.058	0.293	0.039	0.129	0.198	0.519

OLS regressions
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.2: Influences on the dependent variables *MAP* and *Expected Reward Probability* (OLS regressions)

Breaking the Rules: Anticipation of Norm Violation in a Binary-Choice Trust Game

Wolfgang Breuer* Christiane Helduser* Philipp Schade*

Abstract

Studying a binary-choice trust game, we find that trustors who consider trustees to be obligated to reciprocate are ceteris paribus less willing to trust. We argue that norm-orientated trustors anticipate that they will experience negative emotions should trustees violate the obligation to reciprocate. In order to avoid norm violation, trustors lower their willingness to trust. An additional binary-choice trust game, in which subjects act as trustors and subsequently as trustees, reveals that the obligation that trustors assign to the trustee role may be based on social projection.

JEL Classification: C91, D81

Keywords: Trust, Norm, Reciprocity, Reference Point, Laboratory Experiment

* Wolfgang Breuer, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 93539, Fax: +49 241 80 92163, E-Mail: wolfgang.breuer@bfw.rwth-aachen.de.

* Christiane Helduser, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 93561, Fax: +49 241 80 92163, E-Mail: christiane.helduser@bfw.rwth-aachen.de.

* Philipp Schade, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 94774, Fax: +49 241 80 92163, E-Mail: philipp.schade@bfw.rwth-aachen.de (corresponding author).

1 Introduction

A large body of literature emphasizes the influence of norms and obligation on decision making, especially in the context of trust. Interestingly, most studies have focused on the role of the responder, suggesting that trustees may feel an obligation to behave reciprocally when receiving a positive amount (e.g. Ostrom, 2003). These findings are in line with psychological research which postulates the existence of a "norm of reciprocity" (Gouldner, 1960). To our knowledge, there is only one study that explicitly takes into account the influence of obligation on *trustors*' behavior. Buchan, Croson, and Solnick (2008) examine whether trustors feel obligated to trust, i.e. to send money to the trustee, and find that a feeling of obligation increases trustors' likeliness of sending any money. In this paper, we propose an explanation for why obligation feelings may *lower* trustors' willingness to trust, rather than increasing it. As an extension of the previous literature, we do not only examine trustors' own obligation felt, but also their attitudes toward *trustees*' obligation to reciprocate trust. In this context, we also investigate whether trustors act in accordance with these attitudes when taking the trustee role.

2 Related Literature and Hypotheses

As outlined above, it is still unclear whether trustors' behavior is driven by the obligation which trustors assign to the trustee role. In fact, norm-orientated trustors may ex ante evaluate trustees' behavior on the basis of certain norms (e.g. the norm of reciprocity), because norms are not only positive or negative internal valuations for one's own behavior, but also for the behavior of others (e.g. Broom & Selznick, 1963; Williams, 1960). Thus, differences in norm

orientation among trustors may result in different evaluations of trustees' behavior. Obviously, a norm-orientated trustor should be more concerned about non-reciprocal, i.e. norm-violating, behavior of others than a trustor who is not norm-orientated. People who believe in a certain norm may suffer negative emotions or disappointment when this norm is violated by others (Wilson & O'Gorman, 2003). From the trustors' perspective, the simplest way to reduce this "risk" is to lower their willingness to trust. Under this assumption, both the obligation that a trustor feels with respect to her or his own decision and the obligation that she or he assigns to the trustee's decision exert opposite effects on trust. Presumably, both kinds of obligation feelings are positively correlated, because both reflect a more general behavioral pattern of norm orientation. Therefore, identifying the effect of anticipated norm violation on trust requires a simultaneous consideration of trustors' own feelings of obligation and the obligation that trustors assign to the trustee role. We propose the following two hypotheses to cover both kinds of obligation feelings:

H1a: *Trustors who consider trustees to be obligated to reciprocate are less willing to trust.*

H1b: *Trustors who feel obligated to trust are more willing to trust.*

In a second step, we investigate a potential source of the obligation that trustors assign to the trustee role. According to the social projection hypothesis, we expect that trustors consider trustees to be obligated to reciprocate, because the former have internalized the norm of reciprocity and would thus themselves behave reciprocally in the trustee role. Therefore, we expect trustors to act in accordance with this norm when actually taking the trustee role in a repeated trust game:

H2: *Trustors who consider trustees to be obligated to reciprocate are more willing to reciprocate when they act as trustees.*

We refrain from hypothesizing a certain relationship between a subject's own obligation felt in the trustor role and her or his willingness to reward trust in the trustee role, since there might be two counteracting effects. One might expect subjects who do not feel obligated in the trustor role to exhibit a low willingness to reward in the trustee role. However, subjects who do not feel obligated in the trustor role are likely not to trust and might thus be positively surprised when they are trusted in the trustee role, implying a higher willingness to reward.

3 Experimental Design and Procedures

We conducted two experiments in the laboratory of a large German university, using z-Tree (Fischbacher, 2007). In both experiments, subjects first played the binary-choice trust game introduced by Bohnet and Zeckhauser (2004). A particular feature of this game is that a rational trustor's behavior should not be affected by her or his trustworthiness beliefs (Bohnet, Greig, Herrmann, & Zeckhauser, 2008; Bohnet, Herrmann, & Zeckhauser, 2010). The irrelevance of trustworthiness beliefs may be helpful to identify undistorted effects of obligation feelings. Second, subjects' risk preferences were elicited by means of the commonly used multiple price list method (Holt & Laury, 2002). The experiments only differed in one aspect: In the "Single Role" (*SR*) experiment (five sessions with 140 participants), each subject played the binary-choice trust game only once, either as trustor or as trustee, whereas subjects subse-

quently took both roles in the "Both Roles" (*BR*) experiment (four sessions with 116 participants). Recall that subjects playing both roles are required with respect to *H2*. In the *BR* experiment, subjects were informed about the switch of roles only after having completed the first round. Moreover, they got to know that in the second round, they would be paired with a different person than in the first round.

In both experiments, the core of the binary-choice trust game followed the same procedure as in Bohnet and Zeckhauser (2004), including the payoff structure. One point was equal to 0.50 Euros (see Figure 1). First of all, participants were randomly assigned either to the trustor or to the trustee role. After the assignment of roles, trustors reported the probability at which they would prefer to trust rather than to receive the sure payoff, i.e. their *minimum acceptable probability* (*MAP*). Trustors were not informed about the actual reward probability (rp^a) and all payoffs until the very end of the experiment. A reported *MAP* below or equal to the proportion of trustees who reward trust indicates that a trustor is willing to trust ($MAP \leq rp^a$), whereas a reported *MAP* above that proportion indicates the contrary ($MAP > rp^a$). Simultaneously, trustees had to choose between Option 1 (i.e. to reward trust) and Option 2 (i.e. not to reward trust). Having revealed their *MAP*, trustors (Player A) had to state on a four-point scale whether they agreed with the two following statements: "Player B has an obligation to choose Option 1" (variable: *OBL1*) and "I felt obligated to state a lower value for the *MAP*" (variable: *OBL2*). Moreover, trustors were asked to estimate the actual rate of reward (rp^a). This estimation task was incentivized. Note that neither the estimation task nor the obligation questions were part of the experiments conducted by Bohnet and Zeckhauser (2004).

>>> Insert Figure 1 about here <<<

4 Results

To test *H1a* and *H1b*, we consider trustors from both experiments, except for those who were first assigned to the trustee role in the *BR* experiment, i.e. "second round" trustors. The reason behind this is simply that we expect "second round" trustors to identify more with the trustee role than "first round" trustors do. Presumably, closing the empathy gap between the roles may reduce a trustor's cognitive dissonance in the case of norm violation committed by the trustee. As we control for risk aversion, we only consider subjects holding consistent risk preferences.

Since the *MAP* is a fraction, we base our investigation of *H1a* and *H1b* on generalized linear models with a logistic link function (Papke & Wooldridge, 1996). Besides our two main variables of interest, *OBL1* and *OBL2*, we take into account trustors' risk aversion, gender, and their belief about the actual reward probability (variable: *Trustworthiness Belief*) as control variables. The latter was derived from the estimation task. In addition, we checked that multicollinearity is not an issue in all regressions. Model 1 in Table 1 confirms that risk aversion, but not gender and trustworthiness beliefs affect the *MAP*. More importantly, it reveals a negative effect of *OBL2* and a positive effect of *OBL1*. Both findings provide support for *H1a* and *H1b*. Probably as a consequence of increasing empathy, which has been outlined above, the effect of *OBL1* disappears if we include "second mover" trustors in our sample (results not reported). Furthermore, it should be noted that the positive effect of *OBL1* on the *MAP* does not necessarily imply a negative univariate correlation between a subject's willingness to trust as a trustor and her or his willingness to reward trust as a trustee, due to the influence of the

other explanatory variables. Actually, there is no statistically significant correlation between the *MAP* and *Reward* (r = -0.108, p = 0.248).

>>> Insert Table 1 about here <<<

In order to test *H2*, we consider all subjects taking part in the *BR* experiment. As our variable of interest (*Reward*) has two possible outcomes, we run a binary logistic regression including *OBL1* as predictor (see Model 2 in Table 1). We control for *OBL2,* gender, and for whether subjects acted as "first round" or "second round" trustors (variable: *Order*). Model 2 reveals that *OBL1* has a significant positive influence on *Reward*, as predicted by *H2*. Neither *OBL2* nor gender and *Order* play a significant role. We outlined a possible explanation for the insignificance of *OBL2* in Section 2.

5 Conclusion

Our study sheds new light on the role of obligation in the context of trust: Trustors who consider trustees to be obligated to reciprocate are ceteris paribus less willing to trust. Moreover, trustors behave in accordance with their attitudes toward trustees' obligation when taking the trustee role. Our results may be interpreted as the outcome of reference-point-dependent preferences. According to prospect theory (Kahneman & Tversky, 1979), negative deviations from a reference point cause a loss of utility. In the trust game, norms related to reciprocal behavior may constitute such a reference point for trustors: If trustees violate them, trustors suffer a non-monetary loss of utility. According to Alm, McClelland, and Schulze (1999), different social norms might be a potential reason for varying reference points.

Moreover, our findings may provide evidence for the research by Bohnet et al. (2010), who trace back cultural differences in trusting behavior to different reference points for trustworthiness, which we think may reflect differences in social norms across cultures. From an evolutionary perspective, the anticipation of norm violation may help to "stabilize" the overall level of trust in repeated interactions if the first causes a smaller decline in trust ex ante than actual disappointment might eventually do ex post. Additionally, our findings may be linked to the observation that trustworthiness grows continuously over the human life-cycle, whereas the degree of trust only increases in the early years of life (e.g. Sutter & Kocher, 2007). In line with our argumentation, the anticipation of norm violation might hinder trust from increasing at a later age, once people have actually experienced negative emotions resulting from betrayal. However, more research is required to gain a better understanding of both aspects outlined above.

Acknowledgments

We thank Anselm Hüwe, Stefan Kleeschulte, Christine Stibbe, an anonymous referee, as well as seminar participants at RWTH Aachen University for valuable comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alm, J., McClelland, G. H., & Schulze, W. D. (1999). Changing the Social Norm of Tax Compliance by Voting. *Kyklos*, *52*(2), 141–171.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, *98*(1), 294–310.

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). Trust and the Reference Points for Trustworthiness in Gulf and Western Countries. *Quarterly Journal of Economics*, *125*(2), 811–828.

Bohnet, I., & Zeckhauser, R. (2004). Trust, Risk and Betrayal. *Journal of Economic Behavior and Organization*, *55*(4), 467–484.

Broom, L., & Selznick, P. (1963). Sociology (3rd ed.). New York: Harper and Row.

Buchan, N. R., Croson, R. T. A., & Solnick, S. (2008). Trust and Gender: An Examination of Behavior and Beliefs in the Investment Game. *Journal of Economic Behavior and Organization*, *68*(3), 466–476.

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

Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, *25*(2), 161–178.

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, *92*(5), 1644–1655.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, *47*(2), 263–291.

Ostrom, E. (2003). Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation. In E. Ostrom & J. Walker (Eds.), *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research* (pp. 9–79). New York: Sage.

Papke, L. E., & Wooldridge, J. M. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics*, *11*(6), 619–632.

Sutter, M., & Kocher, M. G. (2007). Trust and Trustworthiness Across Different Age Groups. *Games and Economic Behavior*, *59*(2), 364–382.

Williams, R. M. (1960). American Society: A Sociological Interpretation (2nd ed.). New York: Alfred A. Knopf.

Wilson, D. S., & O'Gorman, R. (2003). Emotions and Actions Associated with Norm-Breaking Events. *Human Nature*, *14*(3), 277–304.

Tables

Model	(1)	(2)
Dependent Variable	*MAP*	*Reward*
OBL1	0.198***	0.743***
	(0.071)	(0.269)
OBL2	-0.450***	-0.067
	(0.126)	(0.363)
Trustworthiness Belief	-0.259	
	(0.355)	
Gender (1 = *Male*)	-0.094	0.005
	(0.186)	(0.335)
Risk Aversion	0.159**	
	(0.069)	
Order		0.425
		(0.561)
Constant	0.191	-1.903***
	(0.506)	(0.237)
Pseudo R-squared		0.110
Sessions from *SR*	5	
Sessions from *BR*	4	4
Observations	117	116

(1) Generalized linear model with logistic link function on the *MAP*
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

(2) Binary logistic regression on *Reward* (1 = *Yes*)
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1: Determinants of the *MAP* and *Reward*

Figures

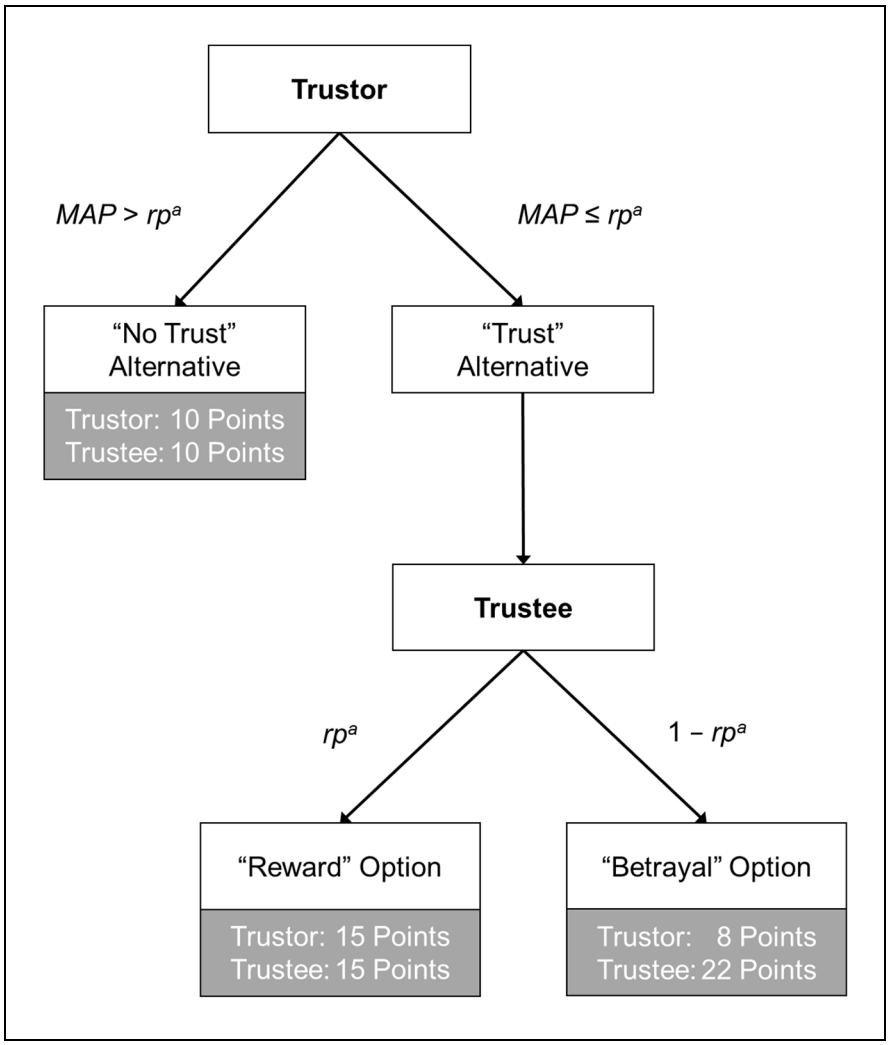

Figure 1: Binary-choice trust game

Once Bitten, Twice Shy: Trust, Trustworthiness, and the Hot-Cold Empathy Gap

Wolfgang Breuer* Philipp Schade*

Abstract

We study a repeated binary-choice trust game in which subjects take the trustor role and then take the trustee role, and vice versa. In line with previous studies, we find that individuals' willingness to trust in the trustor role is positively correlated with their willingness to reward trust, i.e. to act trustworthily, in the trustee role. However, this correlation is moderated by subjects' expectations about the trustworthiness of others. If people hold optimistic trustworthiness expectations, their willingness to trust is positively related to their willingness to reward. The contrary relationship is true for pessimists. We argue that from an individual perspective, the expected utility of trusting depends on the emotions which people expect to experience when being rewarded or betrayed. Optimists tend to overweight positive emotions, whereas pessimists place more emphasis on negative emotions. Therefore, the moderator effect of trustworthiness expectations occurs.

JEL Classification: C91, D81

Keywords: Expectations, Hot-Cold Empathy Gap, Social Preferences, Trust, Trustworthiness

* Wolfgang Breuer, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 93539, Fax: +49 241 80 92163, E-Mail: wolfgang.breuer@bfw.rwth-aachen.de.

* Philipp Schade, RWTH Aachen University, School of Business and Economics, Templergraben 64, 52056 Aachen, Germany. Phone: +49 241 80 94774, Fax: +49 241 80 92163, E-Mail: philipp.schade@bfw.rwth-aachen.de (corresponding author).

1 Introduction

The importance of trust and the willingness to reward trust (i.e. trustworthiness) is widely undisputed among economists. In the field of experimental economics, most studies build on the trust game by Berg, Dickhaut, and McCabe (1995). This game is typically run one-shot, i.e. participants are assigned either to the trustor or to the trustee role. Surprisingly, the question of how trust and trustworthiness are correlated *within* individuals has only aroused little interest up to now. Prominent models which have been set up to explain behavior in social dilemmas imply a negative correlation between trust and trustworthiness (see, e.g., Bolton & Ockenfels, 2000; Dufwenberg & Kirchsteiger, 2004; Fehr & Schmidt, 1999). However, there is empirical evidence challenging this implication: Altmann, Dohmen, and Wibral (2008) run an experiment where participants take the trustor role and then take the trustee role, and vice versa. The authors find a positive intrapersonal correlation between trust and trustworthiness. This finding has been replicated in a sequential-moves prisoner's dilemma (Blanco, Engelmann, & Normann, 2011) and a sequential voluntary contribution game (Gächter, Nosenzo, Renner, & Sefton, 2012).

Thus, the underlying mechanisms of the correlation between trust and trustworthiness are still not fully understood. It seems likely that the positive correlation between trust and trustworthiness is driven by certain (social) preferences, which likewise promote trust and trustworthiness. Altruistic individuals, for instance, may choose to send higher amounts as trustors and to return higher amounts as trustees than non-altruistic ones do, because the former derive utility simply from increasing the payoff of their counterparts. From a more theoretical point of view, however, the preference-based link between trust and trustworthiness

does not necessarily induce a positive correlation between both. To illustrate the ambiguity of the preference-based link, consider a binary-choice trust game, where the trustor decides whether to trust and where the trustee can either reward or betray the trustor's trust.

Let rp^e be the trustor's expected reward probability. First, we abstract from social preferences. Then, her or his utility in the case of trust being rewarded is denoted as U_{reward}, while $U_{betrayal}$ is her or his utility in the case of trust being betrayed. In such a situation, the expected utility of trusting, U_{trust}, is simply

$$U_{trust} = rp^e \cdot U_{reward} + (1 - rp^e) \cdot U_{betrayal}. \tag{1}$$

The expected utility of trusting changes in two aspects if the trustor is assumed to have social preferences:

$$U_{trust} = rp^e \cdot (U_{reward} + \Delta U_{reward}) + (1 - rp^e) \cdot (U_{betrayal} + \Delta U_{betrayal}). \tag{2}$$

On the one hand, the "social" trustor experiences an additional utility increase ($\Delta U_{reward} > 0$) when her or his trust is rewarded. On the other hand, she or he suffers an additional utility decrease ($\Delta U_{betrayal} < 0$) in the case of trust being betrayed. Under the assumption that the existence of social preferences increases an individual's willingness to reward trust (i.e. her or his trustworthiness) in either case, the relative size of ΔU_{reward} and $\Delta U_{betrayal}$ determines whether trust and trustworthiness are positively or negatively correlated within this person. As social preferences simultaneously imply $\Delta U_{reward} > 0$ and $\Delta U_{betrayal} < 0$, their total effect on the relationship between trust and trustworthiness is a priori ambiguous. If

$$rp^e \cdot \Delta U_{reward} + (1 - rp^e) \cdot \Delta U_{betrayal} > 0, \tag{3}$$

social preferences increase one's willingness to trust and induce thus a positive correlation. The contrary is true for $rp^e \cdot \Delta U_{reward} + (1 - rp^e) \cdot \Delta U_{betrayal} < 0$.

The goal of the present paper is to contribute to a better understanding of how trust and trustworthiness are related within individuals if social preferences are assumed. Apparently, according to (3), people's expectations about others' trustworthiness, i.e. their trustworthiness expectations (rp^e), are relevant in this context. In fact, trustworthiness expectations may connect trust and trustworthiness in several ways. The most straightforward effect is that U_{trust} in (1) is an increasing function of rp^e. However, besides this direct effect, there are two other, more indirect ways in which expectations about others' trustworthiness determine the relationship between an individual's own trustworthiness and her or his willingness to trust. These are depicted in Figure 1. One of them is based on a consensus effect (Paths 1 and 2 in Figure 1): Trustworthy people expect others to be similarly trustworthy, and are therefore more willing to trust than people who themselves are not trustworthy (e.g. Altmann et al., 2008; Blanco, Engelmann, Koch, & Normann, 2014; Mullen et al., 1985; Ross, Greene, & House, 1977). In a recent contribution, Breuer, Helduser, and Schade (2017) show that trustors evaluate the outcomes of a binary-choice trust game with respect to their initial expectations. As people are averse to expectations being unfulfilled, they lower their willingness to trust in order to avoid an unexpected loss as a result of betrayal. Note that Breuer et al. (2017) build on social value orientation as a proxy for social preferences rather than on trustworthiness observed in a trust game. The implications of this divergence are discussed in more depth later on. The studies cited above suggest that trustworthiness beliefs *mediate* the intrapersonal relationship between trust and trustworthiness. In the present study, by contrast, we hypothesize that trustworthiness expectations may also constitute a *moderator* (Path 3 in Figure 1). Such a moderator

effect may be driven by trustworthiness expectations (rp^e) increasing one's willingness to trust for given values $\Delta U_{reward} > 0$ and $\Delta U_{betrayal} < 0$, according to (3). As we will outline in more depth, we go one step further by proposing that rp^e may also affect the size of ΔU_{reward} and $\Delta U_{betrayal}$.

>>> Insert Figure 1 about here <<<

The remainder of the present paper is structured as follows. In Section 2, we introduce the binary-choice trust game by Bohnet and Zeckhauser (2004), which is one of the key features of our study, because it eliminates the positive direct effect of trustworthiness beliefs on trust, which has been outlined above, and allows an undistorted view of the hypothesized moderator effect of trustworthiness expectations. This is followed by Section 3, where we develop our hypotheses. Section 4 outlines our experimental design and the respective procedures. In Section 5, we present our results and interpret them with respect to our hypotheses. Section 6 concludes.

2 Minimum Acceptable Probability Design

The Minimum Acceptable probability design by Bohnet and Zeckhauser (2004) is based on a binary-choice trust game (Camerer & Weigelt, 1988; Kreps, 1990). As depicted in Figure 2, the binary-choice trust game has three possible outcomes. If the trustor is not willing to trust ("no trust" alternative), both the trustor and the trustee earn a sure payoff, which we assume to be 10 points for each of them. If the trustor is willing to trust ("trust" alternative), her or his payoff is determined by the trustee. The trustee has to decide whether to reward ("reward"

option) or to betray ("betrayal" option) the trustor. In the case of reward, the trustor and the trustee each receive the same payoff (15 points), which is higher than the first outcome. In the case of betrayal, the trustor's payoff (8 points) is smaller than the sure payoff in the case of not trusting. The trustee's payoff is higher (22 points) in comparison to both the case of not trusting and the case of rewarding.

>>> Insert Figure 2 about here <<<

In the particular framework of Bohnet and Zeckhauser (2004), trustors have to state the probability at which they are indifferent between trusting and receiving the sure payoff, i.e. their *minimum acceptable probability* (*MAP*). The *MAP* represents trustors' switching point beyond which they are willing to take a risk by trusting rather than receiving the sure outcome by not trusting. An *MAP* below or equal to the proportion of trustees who reward trust, i.e. the actual reward probability (rp^a), indicates that trustors are willing to trust ($MAP \leq rp^a$), whereas an *MAP* above that proportion indicates the contrary ($MAP > rp^a$). Consequently, the higher a trustor's *MAP*, the less she or he is willing to trust. Trustors are informed about the actual reward probability only after they have reported their *MAP*.

The *MAP* design offers the important advantage that trustors' willingness to trust, measured in terms of their *MAPs*, should not be directly affected by their trustworthiness expectations (Bohnet, Greig, Herrmann, & Zeckhauser, 2008; Bohnet, Herrmann, & Zeckhauser, 2010). To explain this crucial aspect in a more rigorous way, we take a closer look at the trustors' decision problem. Assume $U_{no\ trust}$ to be the utility derived from choosing the safe outcome and let $U_{reward} > U_{no\ trust}$ as well as $U_{betrayal} < U_{no\ trust}$ denote the utility if a trustor's trust is rewarded or betrayed, respectively. Thus, rational trustors have to balance the sure utility

derived from not trusting ($U_{no\ trust}$) and the expected utility from trusting by choosing their *MAPs* according to the following equation:

$$U_{no\ trust} = MAP \cdot U_{reward} + (1 - MAP) \cdot U_{betrayal}. \tag{4}$$

Setting an MAP^{+} below MAP according to (4), one would accept reward probabilities which lead to expected utilities below the utility of the sure outcome ($U_{no\ trust}$). The contrary is true for $MAP^{+} > MAP$. In this case, one would forgo reward probabilities that generate beneficial expected utilities compared to $U_{no\ trust}$. Since a trustor's MAP represents her or his minimum acceptable reward probability, the MAP should be independent of any considerations about the actual reward probability, i.e. her or his expected reward probability rp^{e}.

3 Hypotheses

Decision making based on expected utilities requires that decision makers know the probabilities of the options at hand and the utilities which they can derive from choosing these options (see, e.g., (1) for the expected utility of trusting in the binary-choice trust game). Economists as well as psychologists typically assume that expected utility incorporates anticipated emotions (e.g. Loewenstein, 2000). In the context of trust, this assumption implies that trustors take into account the emotions associated with trust being rewarded or trust being betrayed when evaluating the expected utility of trusting. More formally speaking, anticipated emotions determine the size of ΔU_{reward} and $\Delta U_{betrayal}$ connected with social preferences.

Psychological research suggests that people's abilities to anticipate future emotions (*arousal*) depend on their current emotional state, which is used as an anchor for prediction.

As a consequence, people in a state of low arousal (cold state) have difficulty predicting future emotions, i.e. what it would be like to be in a state of high arousal (hot state). This phenomenon has become known as the "hot-cold empathy gap" in the psychological literature (e.g. Van Boven & Loewenstein, 2005). Galak and Meyvis (2011) extend this literature by suggesting that the hot-cold empathy gap is related to expectations. More specifically, they argue that people who expect to return to an aversive experience feel higher arousal with respect to this experience. Transferred to the trust game, this line of reasoning implies that people's trustworthiness expectations activate a hot state either with respect to the experience of trust being rewarded or to the experience of trust being betrayed. If people expect their trust to be rewarded, they enter a hot state with respect to the experience of trust being rewarded. The contrary effect takes place if people hold pessimistic trustworthiness expectations. In consequence, optimistic subjects have a tendency to overweight positive emotions, whereas pessimistic trustors place more emphasis on negative emotions.

In formal terms, this implies that the MAP is an implicit function of $\Delta U_{reward} > 0$ and of $\Delta U_{betrayal} < 0$, which, in turn, positively depend on rp^e, i.e. $\partial \Delta U_{reward}/\partial rp^e > 0$ and $\partial \Delta U_{betrayal}/\partial rp^e > 0$. Summarizing, we have

$$\Delta U_{trust}(rp^e) = MAP \cdot \Delta U_{reward}(rp^e) + (1 - MAP) \cdot \Delta U_{betrayal}(rp^e). \tag{5}$$

We can explore the indirect relationship between the MAP and rp^e by considering the partial derivatives, which can be inferred from (5) by means of the implicit function theorem:

$$\frac{\partial MAP}{\partial \Delta U_{reward}} = -\frac{MAP}{\Delta U_{reward} - \Delta U_{betrayal}} < 0 \quad \text{and} \tag{6}$$

$$\frac{\partial MAP}{\partial \Delta U_{betrayal}} = -\frac{1 - MAP}{\Delta U_{reward} - \Delta U_{betrayal}} < 0. \tag{7}$$

From (6), (7), $\partial \Delta U_{reward}/\partial rp^e > 0$ and $\partial \Delta U_{betrayal}/\partial rp^e > 0$, we may conclude that $\partial MAP/\partial rp^e < 0$. Thus, we propose the following hypothesis:

H1: *Optimistic trustworthiness expectations support a positive intrapersonal relationship between trust and trustworthiness, whereas the contrary holds for pessimistic trustworthiness expectations.*

Note that the hypothesized moderator effect of trustworthiness expectations is presumably a distinctive feature of the intrapersonal relationship between trust and trustworthiness, which does not necessarily apply to other types of social behavior. Altruists, for instance, do not necessarily attach negative emotions to trust being betrayed, because altruism is unconditional. In contrast to trustworthiness it does not describe a reaction to a certain behavior of another person. Therefore, it is not even clear that we have $\Delta U_{betrayal} < 0$ and thus $\Delta U_{reward} - \Delta \mathrm{U}_{betrayal} > 0$ in the case of purely altruistic trustors. Moreover, there is no reason to assume that ΔU_{reward} and $\Delta U_{betrayal}$ are functions of rp^e for purely altruistic preferences. As a consequence, trustworthiness expectations do not necessarily moderate the relationship between trust and altruism in the way described by (6). This is also why our study differs significantly from that of Breuer et al. (2017), who investigate the intrapersonal relationship between trust and social value orientation.

In order to test *H1*, we conduct a repeated trust game in which subjects act as trustors and subsequently as trustees (or the other way round). Given that the hot-cold empathy gap operates in the way proposed in *H1*, we expect that in a repeated trust game subjects will tend

to feel higher arousal in the second round than in the first round, irrespective of their trustworthiness expectations and thus irrespective of the direction of their arousal. Therefore, we expect trustors to be more prone to elevated estimations of ΔU_{reward} and $\Delta U_{betrayal}$ when making their "second round" decisions. Our second hypothesis reflects this difference between rounds:

H2: *The moderator effect of trustworthiness expectations proposed in H1 is larger for "second round" trustors than for "first round" trustors.*

4 Experimental Design and Procedures

Our experiment took place in the laboratory of a large German university, using z-Tree (Fischbacher, 2007). Participants were invited via ORSEE (Greiner, 2004). We conducted four identical sessions with 28 to 30 participants ($N = 116$). Each session lasted about 45 minutes. Upon arrival, participants were randomly assigned to a computer. Once all participants were seated, they were informed that they could earn cash depending on their decisions and the decisions made by other participants, that the experiment was held anonymously, and that they would be paid privately at the end of the experiment. The experiment itself consisted of two independent parts and a socio-demographic questionnaire. First, subjects played two rounds of the binary-choice trust game developed by Bohnet and Zeckhauser (2004). Second, subjects' risk preferences were elicited by means of the multiple price-list method of Holt and Laury (2002). Both parts were monetarily incentivized.

Given the relatively high complexity of the binary-choice trust game, the experiment was designed with a strong focus on subjects' comprehension. At the beginning of the trust game part, subjects were provided with written instructions which they might use during the whole experiment. The experimental supervisor, being the same person in all sessions, read the instructions aloud and answered subjects' questions individually. Moreover, subjects were given plenty of time to study the instructions on their own. To ensure further that participants understood the *MAP* framework well, they had to pass a comprehension task in the style of Bohnet and Zeckhauser (2004). To be more precise, subjects had to calculate individual payoffs for trustors and trustees based on hypothetical *MAPs* and reward probabilities. Having answered all questions correctly, the first round of the binary-choice trust game began. Subjects were randomly assigned either to the trustor or to the trustee role. Participants who acted as trustors in the first round were assigned to the trustee role in the second round, and vice versa. The reversal of roles was only announced after the first round. Furthermore, subjects were informed that they would be newly paired in the next round. In this vein, we intended to reduce interdependencies between rounds from the participants' perspective.

The procedure of the two rounds of the binary-choice trust game was as follows. After the random assignment of roles, trustors reported the probability at which they would prefer to trust rather than to receive the sure payoff, i.e. their *minimum acceptable probability* (*MAP*). Then, trustors had to estimate the actual rate of reward (rp^e). The estimation task was incentivized by means of a quadratic scoring rule, which is still a common method to incentivize the elicitation of probability beliefs (see, e.g., Gotthard-Real, 2017; Qiu & Weitzel, 2016). However, the quadratic scoring rule has been subject to criticism, because the revealed beliefs may be biased due to risk aversion. As we only moderately incentivized the estimation task (a fully

correct estimation yielded 1.00 Euros), we do not believe that risk aversion diminishes the validity of the elicited beliefs. Nevertheless, we explicitly address this issue in our regression analyses by controlling for subjects' risk aversion.

While trustors stated their *MAPs* and their estimations of rp^e, trustees had to choose whether to reward or to betray in the case of being trusted (strategy method). Options were neutrally labeled. Having made their decisions, trustors and trustees were randomly paired, and payoffs were calculated. The payoff structure is depicted in Figure 2. One point was equal to 0.50 Euros. An *MAP* below or equal to the actual reward probability rp^a indicated that a trustor was willing to trust, whereas $MAP > rp^a$ indicated the contrary. Note that subjects were not informed about the payoffs from either round until the very end of the experiment.

In the second part of the experiment, subjects' risk aversion was elicited by means of the multiple price list method of Holt and Laury (2002). The multiple price list typically contains ten rows, each representing a choice between a "sure" and a "risky" gamble. After subjects had completed the multiple price list, one of their ten decisions was picked at random for remuneration. Subjects were informed about this procedure in advance. The result of the random draw was disclosed to participants only at the very end of the experiment.

5 Results

5.1 Descriptive Statistics

Across all four sessions conducted, 116 subjects took part in our experiment (45 females and 68 males). We exclude three of them (two females and one male) from further consideration, because they had great difficulty completing the comprehension test, which took place at the

beginning of the experiment. Among the remaining 113 subjects, trustors' mean *MAP* was 59.3 % ($SD = 0.255$). "Second round" trustors stated lower *MAPs* ($M = 0.542$, $SD = 0.291$) than "first round" trustors did ($M = 0.645$, $SD = 0.201$). According to a Mann-Whitney test, the difference is not statistically significant ($Z = -1.572$, $p = 0.116$). The overall actual reward probability rp^a, computed across all sessions, was 43.4 %. The corresponding statistics are depicted in Table 1.

>>> Insert Table 1 about here <<<

According to the *estimation task*, trustors expected on average 48.8 % ($SD = 0.255$) of the trustees to be willing to reward trust. Thus, trustors only slightly overestimated the actual reward probability ($rp^a = 43.4$ %). Referring to their trustworthiness expectations, "second round" trustors were slightly more optimistic ($M = 50.7$ %, $SD = 0.245$) than "first round" trustors were ($M = 47.0$ %, $SD = 0.264$). However, a Mann-Whitney test does not indicate a statistically significant difference ($Z = -0.728$, $p = 0.469$). The corresponding statistics are depicted in Table 2.

>>> Insert Table 2 about here <<<

5.2 Hypotheses Testing

In the following, we examine the interplay between trust, trustworthiness, and trustworthiness expectations. We measure subjects' trustworthiness by whether they chose to reward as trustees (variable: *Reward*). Trustors' trustworthiness expectations are derived from the estimation task (variable: *Trustworthiness Expectations*, identical to rp^e). Higher values of *Trustworthiness Expectations* refer to higher levels of optimism. Trust is measured in terms of the *MAP*

stated in the trustor role (variable: *MAP*). A higher *MAP* implies a lower willingness to trust, and vice versa. Since the *MAP* is a fraction, we base our entire analysis on generalized linear models (GLMs) with logistic link functions (Papke & Wooldridge, 1996). However, we obtain quite similar results by using OLS regressions rather than GLMs. The respective regressions can be found in the Appendix. We ensure that multicollinearity is not an issue in all our models.

To gain first insights, we study the unconditional relationship between the *MAP* and *Reward* within subjects (Model 1 in Table 3). Model 1 includes our two main variables of interest, *Reward* and *Trustworthiness Expectations*, and a dummy variable for whether trustors acted as "first round" trustors or "second round" trustors (variable: *First Round*). Moreover, we control for subjects' gender. Model 1 reveals a negative relationship between *MAP* and *Reward* (i.e. a positive relationship between trust and trustworthiness), which confirms previous research (e.g. Altmann et al., 2008). The variable *Gender* is not significant. It is important to note that the negative effect of *Reward* does not necessarily conflict with *H1*, although *H1* predicts a *positive* effect of *Reward* on the *MAP* for pessimistic trustworthiness expectations. The underlying reason for the negative effect of *Reward* may be that our subject sample is dominated by subjects being prone to a consensus effect (Mullen et al., 1985; Ross et al., 1977), which induces a positive correlation between *Reward* and *Trustworthiness Expectations* and thus a negative correlation between *Reward* and the *MAP* due to the hot-cold empathy gap. By implication, there may be relatively few subjects who themselves are trustworthy (untrustworthy) but pessimistic (optimistic) concerning the trustworthiness of others. Our data provides support for the prevalence of a consensus effect. Among the trustworthy subjects, a

majority (77.6 %) expected the reward probability to be higher than 0.5 (optimistic trustworthiness expectations). A similar pattern applies to the untrustworthy subjects: 70.3 % of them expected the reward probability to be equal to 0.5 or smaller (pessimistic trustworthiness expectations).

A second result that can be directly derived from Model 1 is that "second round" trustors exhibit significantly higher levels of trust than "first round" trustors do (positive effect of *First Round* on the *MAP*). This finding points to another facet of the hot-cold empathy gap, because acting as a "second round" trustor certainly fosters people's perspective-taking abilities. If "second round" trustors identify more with trustees than "first round" trustors do, it is natural to assume that the former tend to state lower *MAPs*.

>>> Insert Table 3 about here <<<

In the second step of our analysis, we extend the basic model by including the interaction between *Reward* and *Trustworthiness Expectations* (Model 2). The interaction term *Reward* × *Trustworthiness Expectations* is statistically significant. Moreover, the conditional effect of *Reward* on the *MAP* turns out to be significantly positive, which indicates a negative relationship between trust and trustworthiness for extreme pessimism. For extreme optimism, however, the effect of *Reward* becomes negative. We examine the interaction between *Reward* and *Trustworthiness Expectations* more closely by means of the predicted mean *MAPs*, which are depicted for several levels of *Trustworthiness Expectations* in Figure 3. All variables other than *MAP*, *Reward*, and *Trustworthiness Expectations* are assumed to be at their mean. As the substantial difference of the slopes illustrates, *Reward* and *Trustworthiness Expectations* interact in the hypothesized way: Optimistic trustworthiness expectations are associated with

a positive intrapersonal correlation between trust and trustworthiness, whereas the contrary holds for pessimistic trustworthiness expectations.

In the third step, we examine whether the interaction between *Reward* and *Trustworthiness Expectations* differs between "first round" trustors and "second round" trustors. For this purpose, we split our sample with respect to the variable *First Round* and re-conduct Model 2 for each subsample. In this manner, we obtain two separate models for "first round" trustors (Model 3) and "second round" trustors (Model 4). The regression results show that the interaction effect is significant for "second round" trustors only. The coefficients of the interaction term differ significantly between "first round" trustors and "second round" trustors ($\chi^2 = 8.32$, $p < 0.05$).

To complete our analysis, we repeat the previous steps by including subjects' risk aversion as an additional control variable in the respective regressions. In the context of trust, several studies have pointed to the role of risk aversion, suggesting that risk averse individuals are less willing to trust than risk neutral or risk loving ones, because putting trust in another person implies taking the risk of being betrayed (see Trautmann & Vieider, 2012 for an overview). The variable *Risk Aversion* refers to the number of "safe" options chosen in the multiple price list section of the experiment (Holt & Laury, 2002). Controlling for risk aversion requires excluding three subjects holding inconsistent risk preferences from further analysis. The respective results, which can be found in Table 4, confirm our previous findings. First, *Reward* and *Trust* are positively correlated (Model 1 in Table 4). Note that the effect of *First Round* is not statistically significant when using OLS regressions rather than GLMs (see Model 1 in Table A.2 in the Appendix). Second, *Reward* interacts with *Trustworthiness Expectations* in the way hypothesized in *H1*. Third, the interaction effect is more pronounced for "second round"

trustors than for "first round" trustors ($\chi^2 = 10.23$, $p < 0.01$). In all four models, *Risk Aversion* has a positive impact on the *MAP*, i.e. a negative impact on trust. However, the effects lack statistical significance. In this aspect, our findings correspond to earlier studies (e.g. Eckel & Wilson, 2004). Last but not least, gender does not play a significant role in any of the models.

>>> Insert Table 4 about here <<<

6 Conclusion

In the present study, we shed new light on the intrapersonal interdependencies between trust, trustworthiness, and trustworthiness expectations. Our findings underline that the relationship between trust and trustworthiness is non-trivial and that emotional perspective taking plays a key role in this context. In line with previous studies, we find that subjects' willingness to trust as a trustor is positively correlated with their willingness to reciprocate in the trustee role. However, this correlation is moderated by participants' trustworthiness expectations. If subjects hold optimistic expectations, their willingness to trust is positively related to their willingness to reward. The contrary relationship is true for pessimistic subjects. We argue that from an individual perspective the expected utility of trusting depends on the emotions which people expect to experience when being rewarded or betrayed. Optimistic subjects have a tendency to overweight positive emotions, whereas pessimistic trustors place more emphasis on negative emotions. As a consequence, the moderator effect of trustworthiness expectations occurs.

A better understanding of how trust and trustworthiness are related within individuals is highly important for practitioners, because outside the laboratory the roles of trustors and trustees typically overlap: A person who has to choose whether to place trust in somebody else on one day may be the trusted party on the other day. Especially policy makers might want to know how measures taken to support trust affect trustworthiness, and the other way round. In line with the proverb "Once bitten, twice shy", a person who has been a victim of betrayal once does not only expect the risk of future betrayal to be higher than others do, but also the pain accompanying this experience to be more intense. In this aspect, our findings provide additional evidence for the intuition that extremely negative or positive expectations increase the amplitude of macroeconomic cycles: Pessimism intensifies crises by reducing trustworthy market participants' willingness to trust, whereas optimism fosters booms by promoting a positive relationship between trust and trustworthiness. This is also why promoting trust and trustworthiness is more costly in times when pessimistic expectations dominate.

Acknowledgments

We thank Christiane Helduser, Stefan Kleeschulte, Theresa Schade, Christine Stibbe, as well as seminar participants at RWTH Aachen University for valuable comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Altmann, S., Dohmen, T., & Wibral, M. (2008). Do the Reciprocal Trust Less? *Economics Letters*, *99*(3), 454–457.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *10*(1), 122–142.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2014). Preferences and Beliefs in a Sequential Social Dilemma: A Within-Subjects Analysis. *Games and Economic Behavior*, *87*, 122–135.

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A Within-Subject Analysis of Other-Regarding Preferences. *Games and Economic Behavior*, *72*(2), 321–338.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, *98*(1), 294–310.

Bohnet, I., Herrmann, B., & Zeckhauser, R. (2010). Trust and the Reference Points for Trustworthiness in Gulf and Western Countries. *Quarterly Journal of Economics*, *125*(2), 811–828.

Bohnet, I., & Zeckhauser, R. (2004). Trust, Risk and Betrayal. *Journal of Economic Behavior and Organization*, *55*(4), 467–484.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 166–193.

Breuer, W., Helduser, C., & Schade, P. (2017). The Higher Your Expectations, the Lower Your Trust: Avoiding the Experience of Unfulfilled Expectations (*Unpublished Manuscript*).

Camerer, C., & Weigelt, K. (1988). Experimental Tests of a Sequential Equilibrium Reputation Model. *Econometrica*, *56*(1), 1–36.

Dufwenberg, M., & Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Eckel, C. C., & Wilson, R. K. (2004). Is Trust a Risky Decision? *Journal of Economic Behavior and Organization*, *55*(4), 447–465.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

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

Gächter, S., Nosenzo, D., Renner, E., & Sefton, M. (2012). Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-by-Example. *Economic Inquiry*, *50*(4), 953–967.

Galak, J., & Meyvis, T. (2011). The Pain Was Greater If It Will Happen Again: The Effect of Anticipated Continuation On Retrospective Discomfort. *Journal of Experimental Psychology: General*, *140*(1), 63–75.

Gotthard-Real, A. (2017). Desirability and Information Processing: An Experimental Study. *Economics Letters*, *152*, 96–99.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In V. Macho & Kremer K. (Eds.), *Forschung und wissenschaftliches Rechnen* (63rd ed., pp. 79–93). Göttingen: Gesellschaft für wissenschaftliche Datenverarbeitung.

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, *92*(5), 1644–1655.

Kreps, D. M. (1990). Corporate Culture and Economic Theory. *Perspectives on Positive Political Economy*, 90–143.

Loewenstein, G. (2000). Emotions in Economic Theory and Economic Behavior. *American Economic Review*, *90*(2), 426–432.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The False Consensus Effect: A Meta-Analysis of 115 Hypothesis Tests. *Journal of Experimental Social Psychology*, *21*(3), 262–283.

Papke, L. E., & Wooldridge, J. M. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics*, *11*(6), 619–632.

Qiu, J., & Weitzel, U. (2016). Experimental Evidence on Valuation with Multiple Priors. *Journal of Risk and Uncertainty*, *53*(1), 55–74.

Ross, L., Greene, D., & House, P. (1977). The "False Consensus Effect": An Egocentric Bias in Social Perception and Attribution Processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301.

Trautmann, S. T., & Vieider, F. M. (2012). Social Influences on Risk Attitudes: Applications in Economics. In S. Roeser, R. Hillerbrand, P. Sandin, & M. Peterson (Eds.), *Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk* (pp. 575–600). Berlin: Springer.

Van Boven, L., & Loewenstein, G. (2005). Empathy Gaps in Emotional Perspective Taking. In B. F. Malle & S. D. Hodges (Eds.), *Other Minds: How Humans Bridge the Divide Between Self and Others* (pp. 284–297). New York: Guilford Press.

Tables

	Minimum Acceptable Probabilities (MAPs)		
	Overall	First Round	Second Round
Mean	0.593	0.645	0.542
Standard Deviation	0.255	0.201	0.291
N	113	56	57

Table 1: Minimum acceptable probabilities stated in the binary-choice trust game

	Trustworthiness Expectations (rp^e)		
	Overall	First Round	Second Round
Mean	0.488	0.470	0.507
Standard Deviation	0.255	0.264	0.245
N	113	56	57

Table 2: Trustworthiness expectations stated in the estimation task

Model	(1)	(2)	(3)	(4)
Dependent Variable	*MAP*	*MAP*	*MAP*	*MAP*
Reward (1 = *Yes*)	-0.296***	0.852*	0.016	1.698**
	(0.050)	(0.452)	(0.446)	(0.715)
Trustworthiness Expectations	0.200	0.762***	0.118	1.202***
	(0.255)	(0.163)	(0.107)	(0.245)
Reward × *Trustworthiness Expectations*		-2.055***	-0.494	-3.551***
		(0.734)	(0.792)	(1.088)
First Round (1 = *Yes*)	0.436**	0.485**		
	(0.219)	(0.208)		
Gender	-0.151	-0.167	0.167	-0.552
	(0.328)	(0.303)	(0.297)	(0.404)
Constant	0.287	0.063	0.604***	0.147
	(0.423)	(0.388)	(0.147)	(0.543)
Observations	113	113	56	57

Generalized linear models with logistic link function
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: Determinants of the *MAP* (without *Risk Aversion*)

Model	(1)	(2)	(3)	(4)
Dependent Variable	*MAP*	*MAP*	*MAP*	*MAP*
Reward (1 = *Yes*)	-0.360***	1.052**	0.159	1.798*
	(0.137)	(0.489)	(0.382)	(0.982)
Trustworthiness Expectations	0.215	0.917***	0.074	1.482***
	(0.168)	(0.207)	(0.163)	(0.331)
Reward × *Trustworthiness Expectations*		-2.558***	-0.787*	-3.981**
		(0.826)	(0.457)	(1.609)
First Round (1 = *Yes*)	0.447*	0.510**		
	(0.235)	(0.223)		
Gender	-0.110	-0.114	0.168	-0.520
	(0.288)	(0.258)	(0.290)	(0.489)
Risk Aversion	0.077	0.104	0.119	0.033
	(0.089)	(0.086)	(0.107)	(0.173)
Constant	-0.178	-0.614	-0.059	-0.137
	(0.551)	(0.485)	(0.646)	(1.586)
Observations	110	110	56	54

Generalized linear models with logistic link function
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Determinants of the *MAP* (including *Risk Aversion*)

Figures

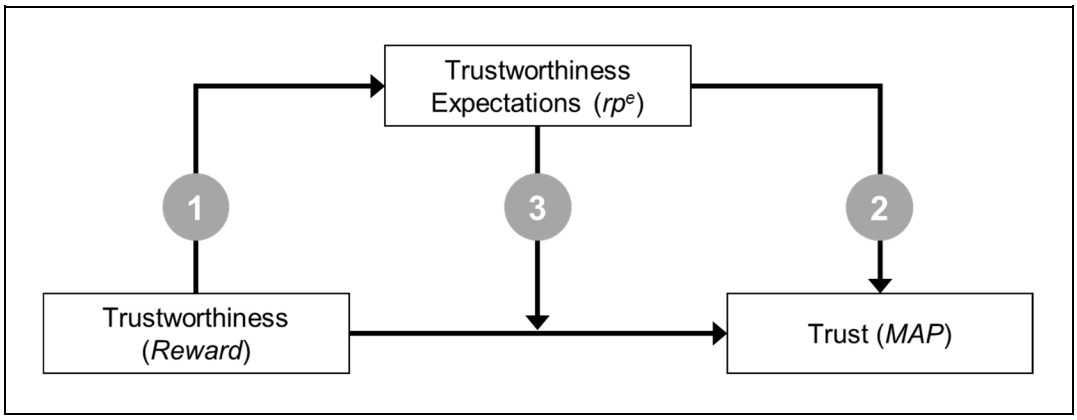

Figure 1: Moderator effect and mediator effect of *Trustworthiness Expectations*

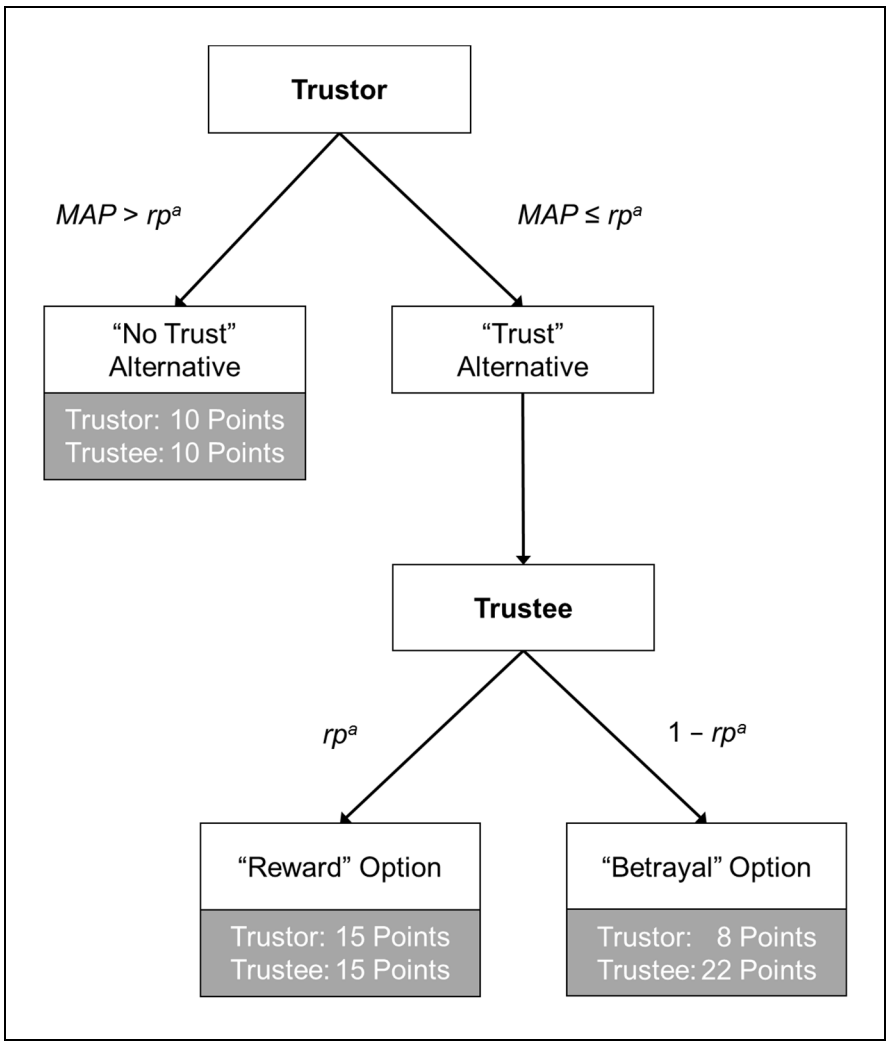

Figure 2: Binary-choice trust game

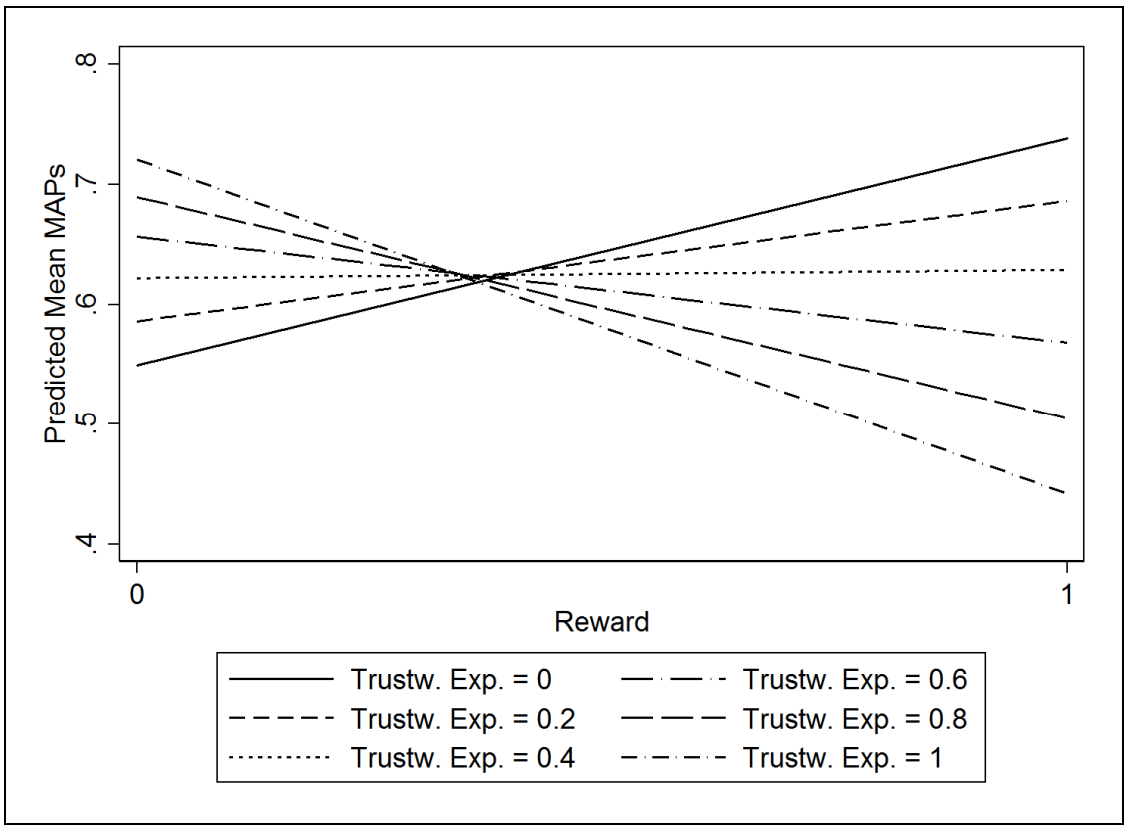

Figure 3: Interaction between *Reward* and *Trustworthiness Expectations*

Appendix

Model	(1)	(2)	(3)	(4)
Dependent Variable	*MAP*	*MAP*	*MAP*	*MAP*
Reward (1 = *Yes*)	-0.296***	0.852*	0.016	1.698**
	(0.050)	(0.452)	(0.446)	(0.715)
Trustworthiness Expectations	0.200	0.762***	0.118	1.202***
	(0.255)	(0.163)	(0.107)	(0.245)
Reward × *Trustworthiness Expectations*		-2.055***	-0.494	-3.551***
		(0.734)	(0.792)	(1.088)
First Round (1 = *Yes*)	0.436**	0.485**		
	(0.219)	(0.208)		
Gender	-0.151	-0.167	0.167	-0.552
	(0.328)	(0.303)	(0.297)	(0.404)
Constant	0.287	0.063	0.604***	0.147
	(0.423)	(0.388)	(0.147)	(0.543)
Observations	113	113	56	57
Adjusted R-squared	0.027	0.052	-0.045	0.065

OLS regressions
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.1: Determinants of the *MAP* (without *Risk Aversion*)

Model	(1)	(2)	(3)	(4)
Dependent Variable	*MAP*	*MAP*	*MAP*	*MAP*
Reward (1 = *Yes*)	-0.086*	0.242	0.036	0.420
	(0.035)	(0.111)	(0.087)	(0.237)
Trustworthiness Expectations	0.052	0.208**	0.018	0.351**
	(0.040)	(0.043)	(0.039)	(0.093)
Reward × *Trustworthiness Expectations*		-0.591*	-0.177	-0.935*
		(0.188)	(0.105)	(0.397)
First Round (1 = *Yes*)	0.106	0.119		
	(0.060)	(0.056)		
Gender	-0.026	-0.027	0.038	-0.122
	(0.069)	(0.061)	(0.068)	(0.118)
Risk Aversion	0.018	0.024	0.027	0.008
	(0.021)	(0.021)	(0.026)	(0.043)
Constant	0.460**	0.363*	0.492*	0.463
	(0.135)	(0.121)	(0.161)	(0.393)
Observations	110	110	56	54
Adjusted R-squared	0.036	0.079	0.003	0.077

OLS regressions
Robust standard errors in parentheses, clustered for experimental sessions
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.2: Determinants of the *MAP* (including *Risk Aversion*)