Engel, Christoph; Zhurakhovska, Lilia

Working Paper

When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives

Preprints of the Max Planck Institute for Research on Collective Goods, No. 2012/16

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Engel, Christoph; Zhurakhovska, Lilia (2013) : When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives, Preprints of the Max Planck Institute for Research on Collective Goods, No. 2012/16, Max Planck Institute for Research on Collective Goods, Bonn

This Version is available at:
https://hdl.handle.net/10419/84981

**Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2012/16**

When is the Risk of Cooperation Worth Taking? The Prisoner's Dilemma as a Game of Multiple Motives

Christoph Engel

Lilia Zhurakhovska

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2012/16

When is the Risk of Cooperation Worth Taking? The Prisoner's Dilemma as a Game of Multiple Motives

Christoph Engel / Lilia Zhurakhovska

August 2012

revised version August 2013

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

When is the Risk of Cooperation Worth Taking?
The Prisoner's Dilemma as a Game of Multiple Motives†

Christoph Engel* / Lilia Zhurakhovska**

Abstract

Both in the field and in the lab, participants frequently cooperate, despite the fact that the situation can be modelled as a simultaneous, symmetric prisoner's dilemma. This experiment manipulates the payoff in case both players defect, and explains the degree of cooperation by a combination of five motives: the size of gains from cooperation, expectations about cooperativeness in the population in question, the degree of risk and loss aversion, and the degree by which a participant is averse to inequity. Information about these motivational forces stems from additional within subjects tests. All five factors are significant only if one controls for all the other motives, which suggests that a prisoner's dilemma is a game jointly characterised by these five motives. The need to control for the remaining explanations seems to be the reason why earlier attempts at explaining choices in the prisoner's dilemma with personality have not been successful.

JEL: C72, C91, D03, H41

Keywords: Prisoner's Dilemma, Efficiency, Belief, Risk Aversion, Loss Aversion, Risky Dictator Game, Conditional Cooperation

† We are most grateful to Hans-Theo Normann for manifold advice and help, and to Sophie Bade and Michael Kurschilgen for comments on an earlier version.

* Max Planck Institute for Research on Collective Goods, corresponding author: engel@coll.mpg.de

** Max Planck Institute for Research on Collective Goods, zhurakhovska@coll.mpg.de

1. Introduction

Elinor Ostrom devoted much of her academic life to social dilemma. She was not sanguine about the potential for conflict and distress, and clearly called for institutional intervention. But she did not share the gloomy picture drawn by Hardin (1968). As one of her titles aptly puts it: she did not believe in social dilemma being a tragedy, but a drama (Ostrom et al. 2002). She perpetually warned against oversimplification, and urged policy makers to first understand the intricate interplay of subtle influences (Ostrom 1990). While she never narrowed her view to a single empirical method, she appreciated the potential of experiments for casting light on the underlying behavioural forces. One of her papers is fairly close to our endeavour (Ahn et al. 2001), so that we believe she would have enjoyed this attempt at showing our gratitude for encouragement and advice for much more than an entire decade.

A host of problems in the game of life have been modelled as prisoner's dilemmas: the quintessential conflict of two prisoners who are independently questioned by the police (Kaminski 2003); a cartel (Bertrand 1883); the conflict of two superpowers who engage in a nuclear arms race (Wiesner and York 1964); global warming (Milinski et al. 2008), to name only a few. If this model is appropriate, the prediction is straightforward. The problem "has technically no solution" (Hardin 1968). Or less colourfully: a prisoner's dilemma is dominance solvable. It has a unique equilibrium in pure strategies. Both (all) players defect from the outset. This even holds if the game is repeated, provided its end is defined (Selten 1978; Rosenthal 1981).

Provided all players maximise their payoffs, and expect all other players to do the same, one needs very little information to make this prediction. Were the other player to cooperate, defection is the best response since this yields the best possible outcome. Were the other player to defect, defection is the best response since this makes sure the player is not the sucker. Consequently, whatever the other player does, this player is always better off defecting. A player need not form beliefs about the other player's action, and there is no need to rely on the Nash equilibrium as the solution concept.

One need not know the concrete payoffs to make this prediction. Without loss of generality, the game may be represented ordinally. Yet intuitively, there is something wrong here. Why would any sensible person consider cooperating when gains from defection are exorbitant? By contrast, is it not reasonable to give cooperation a chance if gains from defection are minuscule, while gains from cooperation are large? Intuitively, cardinality matters. This intuition has been put to the test (starting with Rapoport and Chammah 1965:39), see the lit review below. In the lab, manipulations of cardinality indeed significantly explain behaviour. But why is that?

Cardinality can only matter if participants go beyond maximising their payoffs, in monetary terms. It has been shown that, indeed, the majority of a typical experimental population are willing to cooperate in a dilemma, provided their counterparts cooperate as well (Fischbacher et al. 2001; Fischbacher and Gächter 2010). In an anonymous one-shot game, conditional co-

operators do not know their counterpart's willingness to cooperate. Nonetheless, cooperation in one-shot games or, equivalently, with a stranger design, is substantial (Fehr and Gächter 2000; Keser and van Winden 2000; Brandts and Schram 2001). This implies that conditional cooperators are not only willing to forego the opportunity to exploit their counterparts. They are even accepting uncertainty about their anonymous counterpart's types. Consequently, they must form expectations about the cooperativeness of their counterparts. We expect their willingness to cooperate to be the more pronounced the more they are optimistic about their partner's cooperativeness.

A prisoner's dilemma is a dilemma since the community of players would be better off if all cooperate. This statement holds, irrespective of the cardinality of payoffs. Yet should one not expect more cooperation, the larger gains from cooperation? Indeed, the efficiency motive has been shown to matter in the lab (Engelmann and Strobel 2004). We therefore expect the willingness to cooperate to also grow in gains from cooperation.

Game theorist sometimes speak of greed and fear (e.g. Rapoport 1967). Yet this is only colourful language, merely labelling the incentives to defect, not a statement about mediating emotions. In this paper, we read these labels literally. We propose that participants are more likely to defect in the prisoner's dilemma if they are greedier. We define greed as the degree by which a participant ignores harm she imposes on another participant, for the sake of a larger gain for herself. Greed is the reverse side of the coin of generosity. We apply the concept of inequity aversion (Fehr and Schmidt 1999) to account for generosity. We also propose that participants are more likely to defect if they are more fearful. We use two alternative definitions of fear: risk aversion and loss aversion.

But how can we disentangle these motives? It would be tempting to just manipulate the cardinality of payoffs. Yet no manipulation of parameters can simultaneously isolate all the motives we believe to matter for explaining behaviour in a prisoner's dilemma. We therefore adopt an alternative research strategy. We exploit the fact that greed and fear, as we define them, may be interpreted as personality traits. We therefore can within subjects administer additional tests for these traits, and use them as explanatory variables. We finally elicit beliefs to learn the individual degree of optimism about the cooperativeness of others. All four explanatory factors turn out significant, showing that the prisoner's dilemma is indeed a game of multiple motives, with the motives we expected to matter. The risk of cooperation turns out worth taking if gains from cooperation are substantial, if a participant is sufficiently optimistic that her counterpart will cooperate as well, and if she is not too risk and too loss averse.

On first reading, it may seem surprising that, finally, participants who are more generous in the dictator game are less willing to cooperate in the prisoner's dilemma. The puzzle dissolves if we consider the predominant source of giving in the dictator game. Most participants are not unconditionally generous, but give because they are averse against inequity. They are not altruists, who want to increase other players' payoffs irrespective of their own payoff. Rather they do not only care about absolute, but also about relative payoff. If they outperform anoth-

er participant by too much, they balance things out by giving some of their endowment away. Yet most participants who are averse to advantageous inequity are also averse to disadvantageous inequity, and for most of them being exploited themselves carries more weight than being an exploiter. In the dictator game, the latter risk is absent. Yet the prisoner's dilemma forces participants to choose between the risk of exploiting their random partner (if they defect and the partner cooperates) and being exploited (if they cooperate and their partner defects). Inequity aversion as the common cause induces them to defect in the prisoner's dilemma, the more so the more they cooperate (give) in the dictator game.

Our paper also contributes to the debate about the explanatory power of models of social preferences. These models have been criticized for the fact that they only seem to explain results in the aggregate, not at the individual level. Correlations across games have been found to be low, and usually insignificant (Blanco et al. 2011). That finding seems to put the interpretation of social preferences as traits into question. Now differential psychology for long has established that personality types tend to be conditional. While there is variance across situations, conditional on situation choices tend to be consistent. For consistency it is not necessary that the situation remains the same. It suffices for the situation to be analogous with respect to the trait in question (Ross and Nisbett 1991). Our analysis provides the link that was missing in earlier attempts. While (Blanco et al. 2011) ground their hypotheses in participants' beliefs, they do not elicit beliefs. We do and, controlling for beliefs, find a significant effect of choices in all other games on choices in the prisoner's dilemma.[1]

The remainder of this paper is organised as follows: Section 2 contrasts our approach with the related literature. Section 3 presents the design of the experiment. Section 4 offers our hypotheses. Section 5 is the results section. Section 6 concludes.

2. Related Literature

For decades, prisoner dilemma games have been tested in the lab (for an overview see Colman 1995: 133-160; Sally 1995). A substantial fraction of participants cooperate, and hence violate the theoretical prediction.

Several papers have explored how behaviour reacts to changes in cardinality, of course keeping ordinality such that the game remains a prisoner's dilemma. To organise this literature, we use the labels originally introduced by Rapoport and Chammah (1965:34). In Table 1 R stands for the reward from successful cooperation, S is the sucker payoff, T the temptation payoff, and P signifies the punishment (for defection).

1 For other attempts at explaining choices across games, see the lit review.

	C	D
C	R,R	S,T
D	T,S	P,P

Table 1
Generalized Representation of a Two-Person Two-Action Prisoner's Dilemma

C cooperation, D defection

The closest analogue to our game is Rapoport and Chammah (1965:39). They too vary P, while holding the remaining payoffs constant. They also find that there is the more cooperation the lower P. Yet they only test three different levels of P, while we have a nearly continuous scale. Their test is between subjects, while we test within subjects. Most importantly, they do not run within subjects tests to explain these findings.[2]

Ahn et al. (2001) in a way ask a mirror question. They hold R and P constant, and vary S and T. Therefore cooperation is always equally rewarding, but differently risky. There is most cooperation with high S and low T, and least cooperation with low S and high T. In the interest of neutralising the efficiency motive they manipulate both S and T, which we need not do since we have the additional information from the dictator game and the tests for risk and loss aversion. Major differences result from the fact that they repeat the game (both in a partner and in a stranger design) and that their prisoner's dilemma is preceded by a coordination game. Ahn et al.'s design introduces variance in the difference between the minimum and maximum payoff. One therefore has two options for explaining the results: the absolute difference between S and P or between R and T, or the ratio $(R - P)/(T - S)$. Since we only vary P, in our design the ratio is a linear transformation of P, which is why we can directly work with P. They too do not run within subjects tests to explain their results.

Steele and Tedeschi (1967); Vlaev and Chater (2006) go one step further and directly vary the ratio $(T-S)/(R-P)$. Defection increases the larger this ratio. Finally, as in our design, Schmidt et al. (2001) hold S and T constant, but they simultaneously vary R and P. In regression analysis, both $T - R$ (which they interpret as greed) and $P - S$ (which they interpret as fear) turn out significant. If there is more scope for greed, and if there is more reason for fear, there is less cooperation. Our design differs in that we only vary P, and that we do so quasi continuously. Again no motivational explanations are offered in these papers.

Further papers vary the cardinality of payoffs in more complex games. Rapoport and Eshed-Levy (1989) test participants on three different versions of a five-person step-level public good. These games differ from the one-shot two-person prisoner's dilemma in that defection is not the dominant strategy. If the design gives participants a chance for exploitation, this reduces cooperation more intensely than a design where they only have to fear that the threshold will not be reached. Bruins et al. (1989) vary greed and fear in an eight-person linear public good through manipulating payoff differences. They find about equally strong main effects

2 They do so for different games on p. 45-49.

for fear and for greed. Poppe and Utens (1986) have six participants contribute to or harvest from a common pool resource. Purportedly, the size of the pool is a function of participants' contributions. Actually stated pool size is manipulated by the experimenter. Participants contribute significantly less when the pool is stated to grow, which the authors interpret as evidence that greed has a stronger effect than fear. Kershenbaum and Komorita (1970) have participants simultaneously play two repeated prisoner dilemma games with noisy feedback. In one of these games, they vary the temptation payoff T. If temptation payoffs are unequal, there is much more defection, whether the inequality is to the advantage or to the disadvantage of the player (so that the results do not speak to the greed/fear distinction).

Some related papers pursue different research questions. Van Lange and Liebrand (1990) expose groups of six to prisoner dilemma games with parameters such that in one version there is only reason for fear, in another there is only an opportunity for greed, while the third game combines both. Yet the dependent variable is not choices, but statements about causal attribution. (Simpson 2003aaauthor-year); Kuwabara (2005) compare (between subjects) the prisoner's dilemma with games that arguably only invite fear, greed or the "fear of greed", to show that women are less greedy, but suffer more from the fear of greed.

Previous attempts at explaining choices in the prisoner's dilemma with behaviour in different tests have not been too encouraging. Dolbear and Lave (1966) test participants on three different prisoner dilemma games and do not find any systematic connection with their risk preferences. Swope et al. (2008) combine a prisoner's dilemma with a psychological personality test, which turns out to be insignificant. Boone et al. (1999) do not find a significant effect of psychological measures for locus of control, self-monitoring, aggressiveness and sensation seeking in an anonymous one-shot prisoner's dilemma. Brosig et al. (2007); Blanco et al. (2011) also combine a prisoner's dilemma (and a public-good game) with a dictator game. They find little consistency across games. Yet Kramer et al. (1986); McClintock and Liebrand (1988) significantly explain choices in the prisoner's dilemma with scores from the ring value measure test (Liebrand and McClintock 1988), meant to classify participants' sociality. Boone et al. (2010) also find that prosocials, as classified with this measure, are more likely to cooperate in a prisoner's dilemma, while trust did not have explanatory power.

We believe that the rather sobering earlier findings are due to the fact that the prisoner's dilemma is not a single motive, but a mixed motive game. If these motives moderate each other, or if they interact with each other, one needs a more complete picture of motives to explain choices. Specifically, we believe that conditional cooperation is key to understanding behaviour in the prisoner's dilemma. Therefore the individual degree of optimism should be necessary to understand choices. Our design is meant to provide this more complete picture.

3. Experimental Design

Our dependent variable is choices in one-shot prisoner's dilemma games. We use one-shot games since in repeated games even money maximising agents might cooperate (Kreps et al. 1982). Using the strategy-elicitation method (Selten 1967), participants receive 11 versions of a simultaneous, two-person prisoner's dilemma with T = 10, R = 5, S = 0. We vary P, in the interval [S, R], in the steps as in Table 2.

1	2	3	4	5	6	7	8	9	10	11
0	.05	.2	.5	.8	1.25	1.8	2.45	3.2	4.05	5

Table 2
Prisoner's Dilemma Safety Payoff

This procedure gives us a more fine-grained and a more encompassing dependent variable than in (Rapoport and Chammah 1965:39).

The 11 steps are not equidistant but follow (roughly) $P = x^2 \cdot 0.05$, for $x \in \{0,1,...,10\}$. We choose these parameters to check whether cooperativeness is disproportionately more pronounced if a cooperator has little to loose (since P-S is small) and if gains from cooperation are large (since R-P is large). Note that if S = P, cooperation is "free of charge" – but of course due to T > R, the defection incentive is still present. Conversely, if R = P, there are no gains from cooperation, but a cooperative move still entails the risk of exploitation. We ask all participants to choose between cooperation and defection (neutrally labelled as up and down) in all 11 games. Provided participants choose consistently, i.e. provided they switch from cooperation to defection at a given level of P and they do not switch back with larger P, for each participant we can compress the dependent variable into a single switching point.

Note that we have chosen T = 2R, to make sure that the efficiency motive only matters if both players defect. Motivationally, it only matters if this player expects her counterpart to defect. It then competes with the fear motive. To see this, note that other game theorists interpret P not as a punishment, but as the safety payoff (Straffin 1993:69). Whatever the other player does, if this player plays it safe, she never has less than P. Consequently, the higher P, the more a participant has to lose if she cooperates (P - S). By contrast, we keep gains from defection fixed. They are always given by T – R = R.

Had we only varied P in the prisoner's dilemma, we would not have been able to disentangle motivational forces. If participants are more likely to cooperate when P is small, this could mean that they care about efficiency. Yet it could also follow from the fact that they lose less if they do not get the safety payoff, and hence have less reason for fear; the smaller P, the smaller also P – S. The result could also follow from the fact that greed is not important for these participants, provided greed is addressed to the difference between the safe outcome P and the maximum outcome T; greedy participants would, by contrast, be attracted by the fact that T – P is the larger, the smaller P. Finally, participants might be more willing to cooperate

themselves if P is small since they believe that, with small P, other participants are also more likely to cooperate, so that cooperation is more likely to pay (of course assuming conditional cooperation; otherwise defection maximises profit, whatever the other player does).

To gain information about motivational forces, we run a series of additional tests.[3] At the beginning of the experiment, participants only know that the experiment has more parts, but do not know what these parts are about. Therefore they cannot anticipate later parts of the experiment when making choices in earlier parts. First we ask participants how many of the 24 participants in their session they believe have made the cooperative choice (labelled "up") when P = 2.45 €. We have selected this problem since it is approximately the mean and the median of the support (of the safety payoffs). From pretests, we also expected this to be approximately the empirical mean of switching points in the prisoner's dilemma (which turned out true; the empirical mean is 2.465 €). If a participant guesses the number correctly, we pay 2 €; if the estimate is no further than plus or minus 2 away from the true number, we pay 1 €.

To get information on the individual level of generosity or, conversely, greed we conduct a risky dictator-game (for different versions of the risky dictator game see Bohnet and Zeckhauser 2004; Hong and Bohnet 2007). We ask all subjects to choose between two situations: in situation 1, the proposer and her partner both get 5 €. In situation 2, the proposer has a chance of $0 \leq a \leq 1$ to get 10 €, and a chance of $1-a$ to get 5 €. If the proposer chooses the lottery, the partner gets nothing. The proposer knows this. We vary $a \in [0,1]$, in equal steps of .1. We again ask participants to make their choice for each of the 11 games. At the end of this game, participants are randomly assigned to be dictators or recipients. One problem is randomly selected. Note that, in the prisoner's dilemma, there is both this risk (will the other player cooperate, which is a precondition for receiving 5€?) and a risk of incurring a loss (will the other player defect, which would reduce the payoff to zero?). Our design of the dictator game isolates the former motivational force. Whether the dictator gets a payoff higher than the sure 5€ hinges on the random draw. Yet the dictator can never fall below 5€, whether she is generous with the recipient or not. Note that the expected payoff of the active player is higher in situation 2 whenever a > 0, but the joint payoff of both players is higher in situation 1 as long as a < 1.

To get information on the individual level of fear, we conduct a test for loss aversion. Loss aversion assumes that participants compare payoffs with a reference point. The obvious reference point in the prisoner's dilemma is the safety payoff P. For if she defects, irrespective of the other player's choice, the first player at least receives P. With varying P we vary the reference point. A loss averse individual should switch from cooperation to defection for a lower P

3 One potential drawback of our design results from the fact that participants might feel overwhelmed by the number of tests. We try to tackle the problem by presenting the instructions to each experiment only just before the respective experiment starts. We then have control questions and only start each of the experiments after the participants demonstrate by answering the control questions that they fully understand the instructions. Note that we cannot break down the experiment into several smaller experiments, in which only one or two of the traits are measured, because the very core of our experimental research question is explaining choices in a prisoner's dilemma by the interplay of a series of motives.

than a subject who is not too sensitive to loss aversion. [4] For testing loss aversion, we use the version proposed by Gächter et al. (2007), which is a modification of Fehr and Goette (2007); (for background, see also Rabin (2000a); Rabin and Thaler (2001); Köbberling and Wakker (2005)). In this game, a participant chooses between a safe payoff of zero and a lottery. In the lottery, there is a 50% chance to gain 6 €. In six equal steps of 1 €, we vary the loss in the interval [2 €, 7 €].

Alternatively, fear might result from the fact that participants simply consider cooperation as a risky choice since it may either yield R or only S, while defection gives them a risk-free payoff of (at least) P (on the conceptual bridge between both concepts see Thaler et al. 1997; Rabin 2000b). To be able to compare the two competing interpretations of the fear motive, and to assess whether they interact, we also test participants using the design developed by Holt and Laury (2002). We thus have participants choose between two lotteries. In the first lottery, participants either gain 1.60 € or 2 €. In the second lottery, they either gain .1 € or 3.85 €. We vary the probability of the high gain from 10 % to 100 %, in 10 equal steps.

The experiment was conducted in four separate sessions at the Bonn Econ Lab. Sessions lasted approximately an hour and a half. The sequence was always: Prisoner's Dilemma, Belief Elicitation, Loss Aversion Test, Dictator Game, Risk Aversion Test. There was no feedback at all between any of these decisions.

In total, 96 students participated. The experiment was programmed in z-Tree (Fischbacher 2007). Subjects were recruited with the software ORSEE (Greiner 2004). All subjects were students of various disciplines. Mean age was 24.81 years. 59 (61.46 %) were female. 11 participants (11.46 %) were economics majors.[5] Participants on average earned 10.50 € (4.125 € from the prisoner's dilemma, .25 € from belief elicitation, -.09 € from the test for loss aversion, 2.31 € from the test for risk aversion, and 3.59 € from the dictator game). In the test for loss aversion, losses were possible. To guarantee positive earnings, we announced a minimum payment of 5 €.[6] The minimum payment became effective for 11 participants. For details, the reader is referred to the instructions in the appendix.

All five parts of the experiment were payoff-relevant. In all parts but belief elicitation, one of the problems was randomly selected to be paid out. For the prisoner's dilemma and the dictator game, participants were randomly matched to a partner after they had made all their choices. All random choices were executed by the computer.

---

4 An alternative interpretation is that the reference point is the distance between P and the desired cooperation payoff R. This alternative reference point does not change the prediction about the correlation between loss aversion and the switching point for a given safety payoff P.

5 Their choices did not significantly differ from the choices of the remaining participants, Tobit explaining switching points in the prisoner's dilemma with economics major, p = .245. For the specification of the statistical model see below 5.

6 This method of making sure that payoffs cannot become too small is very unlikely to affect choices given participants were only informed about later parts of the experiment as soon as earlier parts were completed.

4. Hypotheses

If participants hold standard preferences, cooperation is dominated in the prisoner's dilemma if $P > 0$. If $P = 0$, cooperation is still weakly dominated. Hence we would predict defection all over. However, it is known since (Rapoport and Chammah 1965:39) that more cooperation is to be expected with small P. We refine this to

H_1: In a symmetric two-person prisoner's dilemma participants cooperate more the smaller the risk free payoff P.

This hypothesis implies that participants are sensitive to efficiency gains. This presupposes that a substantial fraction of participants is willing to forego the opportunity to exploit their counterparts. While theoretically this could follow from unconditional altruism, in line with Fischbacher et al. (2001); Fischbacher and Gächter (2010) we expect most participants to be only conditionally cooperative. Then their expectations should matter. This leads to

H_2: The more a participant in a symmetric two-person prisoner's dilemma is optimistic about the likelihood of the other participants to cooperate as well, the more she cooperates herself.

In a one-shot game among strangers, a conditional cooperator has no information about the personality of her random counterpart. If she cooperates, she runs the risk of being let down. She may end up with S, where she might at least have secured P. Therefore, if participants are conditional cooperators, risk-aversion should matter. We expect

H_3: The more a participant is risk averse, the less she cooperates in a symmetric two-person **prisoner's** dilemma.

Participants might instead see the safe payoff P as their reference point. Then being let down by a defecting counterpart would imply a loss. This leads to

H_4: The more a participant is loss averse, the less she cooperates in a symmetric two-person **prisoner's** dilemma.

Through our manipulation of P, we also change T-P. The smaller P, the higher is the potential benefit from switching from cooperation to defection and thereby exploiting the other player in case she cooperates. The risky dictator game also empowers the active player to inflict harm on the passive player for the sake of the chance of a larger gain for herself and thereby increasing inequity in their payoffs. This invites

H_5: The **less** a participant gives in the risky dictator game, the less she cooperates in a symmetric two-person prisoner's dilemma.

Arguably, many participants will simultaneously take many or even all of these motives into account when deciding whether to take the risk of cooperation. This leads to our final and decisive hypothesis

H_6: The willingness to cooperate in a one-shot simultaneous prisoner's dilemma is explained by the interplay of gains from cooperation, beliefs about cooperativeness, risk aversion, loss aversion and inequity aversion.

5. Results

a) Gains from Cooperation

We first investigate the effect of increasing the risk free payoff P on cooperation in the prisoner's dilemma. Figure 1 reports the average defection rate across subjects for each realization of P. The figure shows that defection rates monotonically increase in P. This adds to previous findings (e.g. Rapoport and Chammah 1965:39) where only few values of P were analyzed. The visual impression is corroborated by statistical analysis, Table 3.[7] The size of the safety payoff P indeed predicts defection in the respective specification of the prisoner's dilemma.

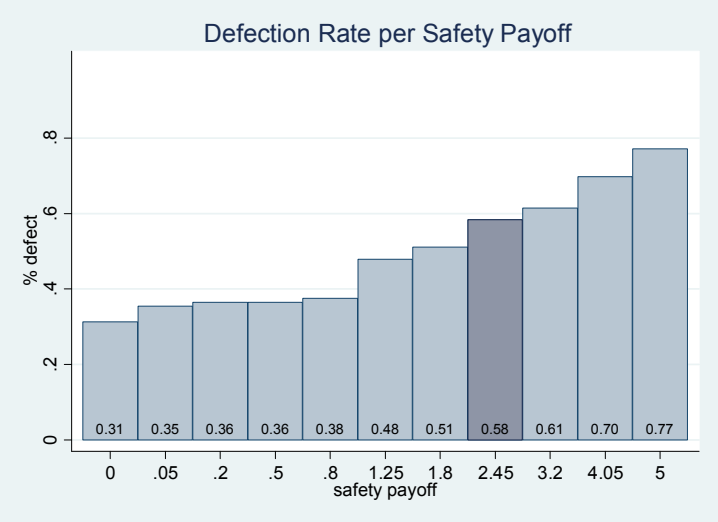


Figure 1
Effect of Payoff in the Case of Mutual Defection on Cooperation

Beliefs have been elicited for the shaded payoff in the case of mutual defection, i.e. for 2.45 €

7 In this regression, we have 11 choices from each participant. We capture the fact that choices are not independent within individuals by an additional individual specific error term. The subsequent Hausman test shows that this procedure is justified, i.e. that the individual specific variance is indeed random.

P	1.195*** (.113)
Cons	-1.757** (.607)
N	1056
p Model	<.001

Table 3
Explaining Defection Rates with Payoff in the Case of Mutual Defection

dv: choices of participants in each of the 11 prison dilemma games
P: payoff in case both players defect
random effects logit, Hausman test insignificant
standard errors in parenthesis
*** p < .001, ** p < .01

We have chosen the steps of P such that we can also see whether cooperation is disproportionately more likely with P at or close to zero. This turns out not to be the case. If we compare defection at P = 0 (where cooperation is only weakly dominated) and at P = .05 (where cooperation is strictly dominated), we do not even find a significant difference (Wilcoxon, N = 96, p = .1025).

In the 11 problems of the prisoner's dilemma, 84 (87.5 %) of our participants behave consistently. They cooperate up to a certain value of the safety payoff P, and defect for higher values.[8] We therefore can compress our dependent variable into switching points.[9]

The switching points turn the original 11 binary variables into one continuous variable. The higher the switching point, the more a participant is cooperative. Figure 2 suggests that we have three types of players. The largest fraction of 31.33 % defect, irrespective of the size of the safety payoff. A not so small minority of 21.69 % always cooperates. The majority (46.98 %) are sensitive to the size of the safety payoff.

8 For the analysis, we (also) drop one data point from the only participant who switches in the opposite direction. This participant cooperates with large P, and defects with small P.

9 Of course, this implies that we cannot use the data from participants who were inconsistent in the prisoner's dilemma.

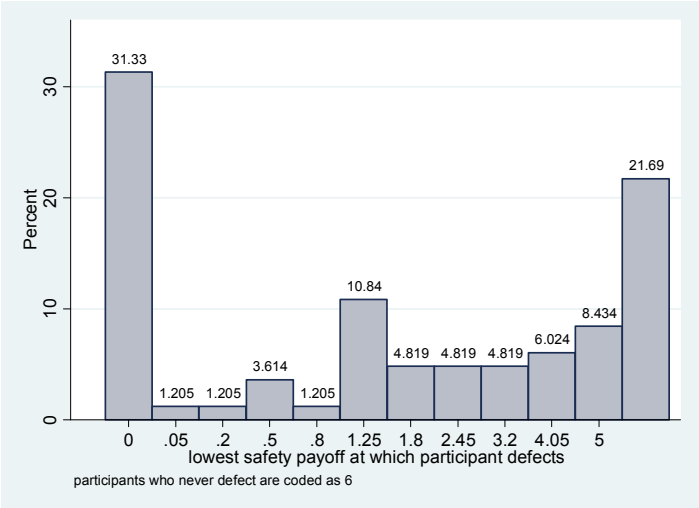


Figure 2
Switching Points in the Prisoner's Dilemma

Our first explanatory variable are gains from cooperation. If P = 0, and if both players defect, they individually and jointly have 0, while they would have had R = 5 for each of them, and 2R = 10 for both, had both of them cooperated. Hence in this game, the total gain from cooperation is 10. By the same token, we can calculate the gain from cooperation for any other game. When they make choices, participants do not know which game will eventually be payoff-relevant. At their individual switching point, the expected value from cooperation is therefore given by the sum of gains from all games up to this point, divided by 11. They range from 0 to 6.49.

Since we have many switching points at both extremes, for analysing this dependent variable, a Tobit model is appropriate. Note that, in this regression, we do not explain a certain value of P by the efficiency gain involved (which would be circular). Rather we explain *the decision* to switch at a certain point by the expected gains from cooperation. The switching points are a condensed way of expressing choices in the individual prisoner's dilemma games, not a measure for the safety payoff. The functional equivalent of this estimation strategy would be a panel model that separately analyses all 11 choices per individual, and explains the decision to defect by the respective size of P. We prefer analysing switching points since our design has invited participants to simultaneously develop a strategy for the entire parameter space.[10]

Gains from cooperation have a highly significant positive effect on switching points in the prisoner's dilemma. The effect from choices in individual problems (Table 3) thus translates into an effect at the individual level, despite the fact that many participants are at both extremes, and therefore not sensitive to the size of gains from cooperation. We thus support $\mathbf{H_1}$, both when working with mean defection per problem and with switching points.

10 This estimation strategy also has a technical advantage. We need not estimate a panel since, for each participant, we only use a single data point, i.e. the switching point. Huber-White standard errors cater for potential heteroskedasticity.

Gains from Cooperation	1.676*** (.290)
Cons	-5.707** (1.696)
N	83
Left Censored	26
Right Censored	18
p Model	<.001
pseudo R^2	.4150

Table 4
Explaining Switching Points With Gains From Cooperation

dv: lowest safety payoff at which participant defects
gains from cooperation: when switching to defection at a certain point
Tobit, robust standard errorsstandard errors in parenthesis
*** p < .001, ** p < .01

b) Optimism

To measure optimism, we elicit beliefs. We ask our participants how many of the 24 participants of their session they believe to have chosen the cooperative move.[11] We do so for the problem with P = 2.45. For our regressions, we translate this estimate into a percentage. As a group, our participants are very well calibrated (for similar results cf. Surowiecki 2004). Actually, 58 % cooperate and the mean belief is 59.81 %. Unsurprisingly, as Figure 3 demonstrates, beliefs are dispersed over the entire range.

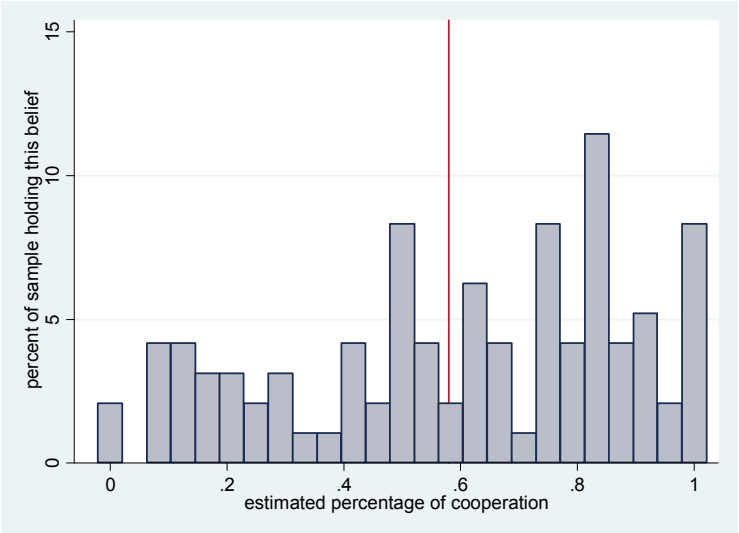


Figure 3
Beliefs

estimated percentage of cooperation at P = 2.45€
red line: true percentage of cooperators

11 As is standard in experimental economics, we have elicited beliefs after the main experiment. With the opposite order, we would have contaminated our dependent variable. Given the order, we cannot completely rule out that stated beliefs are influenced by individual choices. Yet we deem such a consensus effect to be very unlikely. Beliefs are separately incentivized. The object of the belief is cooperativeness in the entire lab, while only one partner matters for payoff. Finally, we only elicit beliefs for one of 11 problems, while participants have made choices for all 11 problems.

Optimism is indeed important for explaining switching points in the prisoner's dilemma, Table 5. This holds if we exclusively explain the switching point in the prisoner's dilemma with optimism (model 1). Yet if we control for gains from cooperation, instead of little more than 10% we explain 45% of the variance (i.e. the pseudo R^2 goes up by almost 35%, model 2). In both specifications, beliefs have a highly significant positive effect on the location of the individual switching point. This supports $\mathbf{H_2}$. Note that this is not what one expects from selfish players. The more they are optimistic that others cooperate, the more, not the less, they should be drawn towards defection, hoping that they will not only gain P but T.

	model 1	model 2
Gains from Cooperation		1.371*** (.236)
Optimism	8.816*** (1.612)	2.885*** (.486)
Cons	-3.039*** (1.042)	-5.914*** (1.539)
N	83	83
Left Censored	26	26
Right Censored	18	18
p Model	<.001	<.001
pseudo R^2	.1086	.4554

Table 5
Explaining Switching Points With Optimism

dv: lowest safety payoff at which participant defects
gains from cooperation: when switching to defection at a certain point
optimism: estimated percentage of cooperation
Tobit, robust standard errors
standard errors in parenthesis
*** p < .001

c) Loss Aversion

In the test for loss aversion, almost all participants were consistent. 94 of 96 (97.92 %) accepted the lottery as long as the potential loss did not exceed a certain amount, and they rejected all lotteries with a higher loss. Figure 4 presents choices in individual problems, and the switching points. The picture is typical for this test. Most participants reject the lottery if the loss is either 4 € or 5 €.

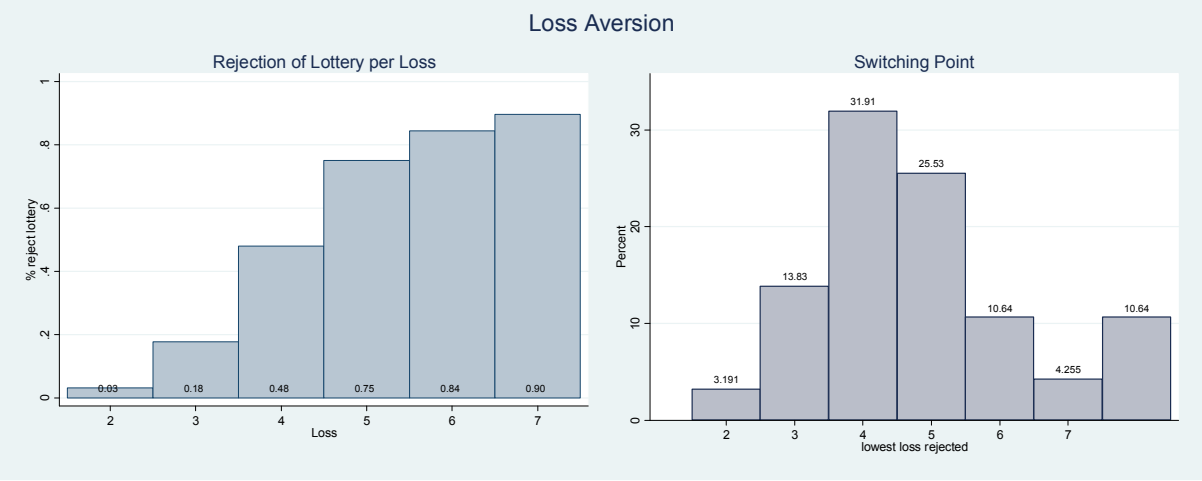


Figure 4
Loss Aversion

right panel: participants who accept all lotteries are coded as 8

If a participant rejects only the last lottery, she is indifferent between receiving 0 with certainty, and both winning and losing 6 € with probability .5. Such a participant is not loss-averse at all. Using this, and in keeping with the standard in the literature (Gächter et al. 2007), we transform the switching point in this test into the λ- value of prospect theory,[12] from which the concept of loss aversion is derived, according to

$$\lambda = 6/switchingp \quad (1)$$

If a participant is loss neutral she has $\lambda = 1$. λ- values range from .75 to 3.[13]

If we try to explain switching points in the prisoner's dilemma with only loss aversion, the coefficient is far from significant (model 1, p = .4803). If we control for gains from cooperation and optimism, we find a weakly significant effect of loss aversion (model 2, p = .073). It surprisingly is positive. We will come back to this effect.

12 If a participant accepts all lotteries, we code the switching point as 8, leading to $\lambda = .75$.

13 In this and the following regressions, results look very similar if we do not transform variables, but directly work with the switching points from the tests for loss aversion, risk aversion, and the risky dictator game. We prefer the transformed values since they are better in line with the underlying behavioural theory.

	model 1	model 2
Gains from Cooperation		1.541*** (.222)
Optimism		2.828*** (.520)
λ	.849 (1.197)	.871^{+} (.479)
Cons	.922 (1.817)	-8.121*** (1.364)
N	83	83
Left Censored	26	26
Right Censored	18	18
p Model	.4803	<.001
pseudo R^2	.0018	.4671

Table 6
Explaining Switching Points With Loss Aversion

dv: lowest safety payoff at which participant defects
gains from cooperation: when switching to defection at a certain point
optimism: estimated percentage of cooperation
λ : degree of loss aversion
Tobit, robust standard errors
standard errors in parenthesis
*** p < .001, $^{+}$ p < .1

d) Risk Aversion

In the test for risk aversion, 92 of 96 participants (95.83 %) are consistent. Choices are as in Figure 5. Few participants are risk-seeking. In their majority, as one would have expected, most participants are risk-averse, yet not extremely so.

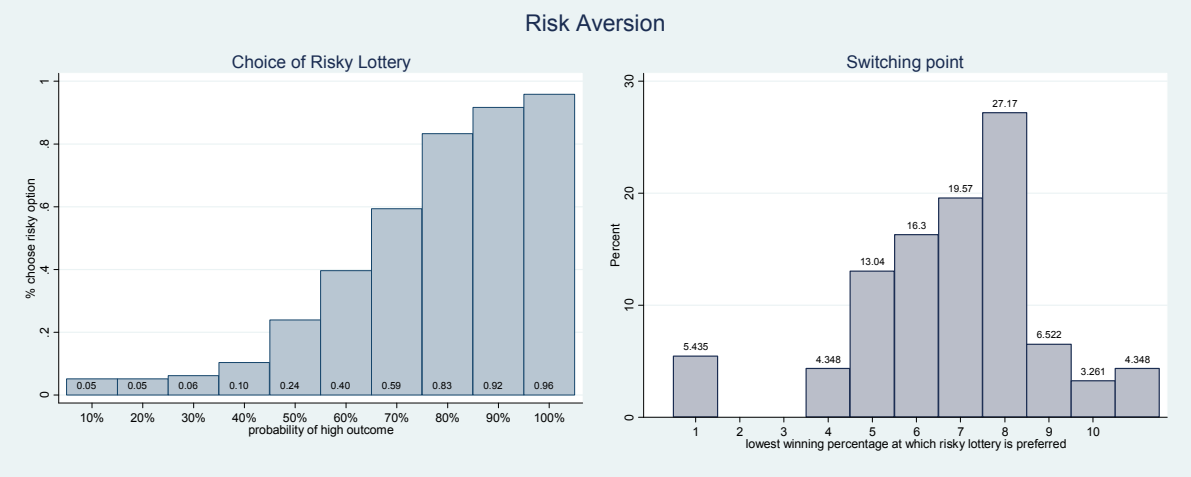


Figure 5
Risk Aversion

right panel: participants who always choose the safe lottery are coded as 11

Following (Holt and Laury 2002:1649), we translate switching points into scores of relative risk aversion, using the following transformation:

$$p*3.85^{1-r}+(1-p)*.1^{1-r}-p*2^{1-r}-(1-p)*1.6^{1-r}=0 \quad (2)$$

At the switching point, the participant is indifferent between both gambles. The transformation rule assumes constant relative risk aversion. Inserting the probability of the gain at the participant's switching point, and solving for r, we generate our measure of relative risk aversion. $r=0$ indicates risk neutrality. Negative values stand for risk seeking behaviour, positive values for risk averse behaviour. In our sample, relative risk aversion ranges from – 1.71 until 1.37, with mean .324.

	model 1	model 2	model 3
Gains from Cooperation		1.518*** (.239)	1.875*** (.138)
Optimism		2.796*** (.502)	2.410*** (.391)
λ		.747^{+} (.445)	-.389 (.271)
Rel.Risk Aversion	.617 (.968)	.160 (.399)	-3.865*** (.522)
λ*Rel.Risk Aversion			3.156*** (.340)
Cons	1.816* (.689)	-7.870*** (1.532)	-8.529*** (.965)
N	81	81	81
Left Censored	26	26	26
Right Censored	17	17	17
p Model	.5253	<.001	<.001
pseudo R^2	.0019	.4668	.5623

Table 7
Explaining Switching Points With Risk Aversion

dv: lowest safety payoff at which participant defects
gains from cooperation: when switching to defection at a certain point
optimism: estimated percentage of cooperation
λ : degree of loss aversion
rel.risk aversion: degree of relative risk aversion
Tobit, robust standard errors
standard errors in parenthesis
*** p < .001, * p < .05, $^{+}$ p < .1

If we try to explain switching points in the prisoner's dilemma with just relative risk aversion,[14] we do not find a significant result, Table 7 model 1. Risk aversion remains insignificant if we add gains from cooperation, optimism and loss aversion as control variables (model 2). It becomes significant though if we also add the interaction with the loss aversion parameter λ. Then risk aversion has the expected negative sign: the more a participant is risk averse, the smaller the value of P at which she switches to defection. Through the interaction term, the effect becomes much smaller, and may even swap signs, if a participant is also pronounc-

14 Compared to the previous regressions, we loose 2 observations due to participants who were inconsistent on the test for risk aversion (but were consistent on the prisoner's dilemma and the remaining tests).

edly loss averse. Further note that, in this model, despite being insignificant, loss aversion also has the expected negative sign. With this qualification, we support $\mathbf{H_3}$. Note the large gain in explained variance resulting from the interaction term. It goes up by almost 10 %. This is due to the fact that, in this sample, risk and loss aversion are only weakly correlated (r = .3452, p = .0007). Many participants score high in one dimension and low in the other. The positive interaction term filters those out who score similarly in both dimensions. For such participants, risk aversion and loss aversion measure a positively correlated motive. The interaction term ensures that the related motive does not count twice. The positive interaction between risk and loss aversion has previously been documented (see Novemsky and Kahneman 2005)).

e) Inequity Aversion

The risky dictator game finally provides us with a measure for one aspect of inequity aversion, the aversion against advantageous inequity. It is sometimes also referred to as generosity, and its absence is referred to as greed. In the risky dictator game, only a single participant is inconsistent. All others leave 5 € to the recipient up to a certain probability of winning 10 €, and they choose the lottery for all higher winning probabilities. Figure 6 shows that we seem to have two groups of participants. Freeriders would either be willing to give the recipient 5 € if this costs them nothing, or they would even choose the selfish option in this case. The remaining participants are willing to decline a small chance to gain 10 €, but give in to temptation at some point.

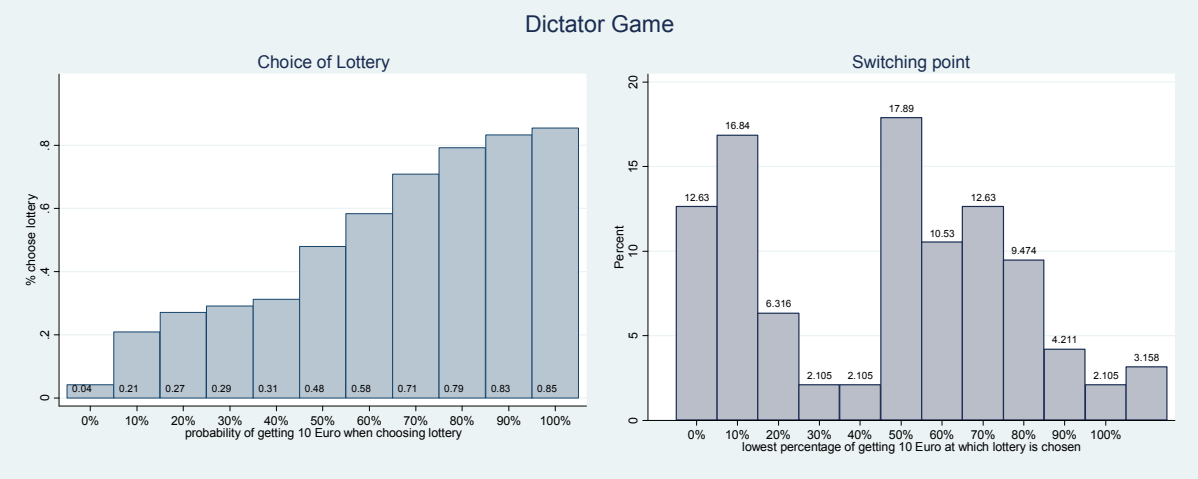


Figure 6
Risky Dictator Game

right panel: participants who never choose the lottery are coded as 110%

In the regressions, we do not directly work with these switching points. We transform them into values of the Fehr and Schmidt (1999) model. For two-player games, (Fehr and Schmidt 1999) propose

$$u_i(\pi_i,\pi_j)=\pi_i-\alpha\max\{\pi_j-\pi_i,0\}-\beta\max\{\pi_i-\pi_j,0\} \quad (3)$$

where π_i is this player's payoff, π_j is the other player's payoff, α is the weight attached to disadvantageous inequity (if any), and β is the weight attached to advantageous inequity (if any). Below the switching point, a participant considers the disutility from leaving nothing to the recipient to be higher than the expected value of the lottery. The participant thus chooses the outcome with the equal distribution as long as $(1-\beta)(5r+5)<5$, where r is the highest probability of winning 10 € at which this participant still gives 5 € to both. Solving for β, we have[15]

$$\beta=\frac{r}{r+1} \quad (4)$$

Note that not calculating an α-value is fully in keeping with the Fehr/Schmidt model. In our game, a dictator can never suffer from disadvantageous inequity. In our sample, β ranges from 0 to .524, with mean .277.

	model 1	model 2
Gains from Cooperation		1.794*** (.124)
Optimism		2.440*** (.317)
λ		-.579* (.249)
Rel.Risk Aversion		-3.401*** (.425)
λ *Rel.Risk Aversion		2.841*** (.292)
β	-1.916 (3.628)	-3.070** (.876)
Cons	2.645* (1.287)	-6.906*** (.943)
N	83	81
Left Censored	26	26
Right Censored	18	17
p Model	.5988	<.001
pseudo R^2	.0011	.6077

Table 8
Explaining Switching Points with Inequity Aversion

dv: lowest safety payoff at which participant defects
gains from cooperation: when switching to defection at a certain point
optimism: estimated percentage of cooperation
λ : degree of loss aversion
rel.risk aversion: degree of relative risk aversion
β: degree of aversion against advantageous inequity
Tobit, robust standard errors
standard errors in parenthesis
*** p < .001, ** p < .01, * p < .05

15 If a participant always gives the recipient 5 €, we code the switching point as 1.1, leading to $\beta=.5238$.

Once more, if we try to explain switching points in the prisoner's dilemma exclusively with inequity aversion, we do not find a significant result (Table 8 model 1). By contrast if we add β to the complete regression, all regressors are significant at conventional levels, including λ, which was still insignificant in Table 7, i.e. when not controlling for inequity aversion. Now both measures of risk and loss aversion have the expected negative sign. Their interaction measures to which degree both measures neutralise each other. We now also support hypothesis $\mathbf{H_4}$ regarding loss aversion. The model has a very good fit. It explains more than 60 % of the variance, which is again more than in the best model of Table 7, where we explained 56% of the variance.

We do, however, have a surprising effect: the later participants give in to temptation in the risky dictator game and chose the lottery that gives the passive player 0, i.e. the higher their β, the earlier they defect in the prisoner's dilemma. It seems as if, generosity in the dictator game made subjects more selfish in the prisoner's dilemma. Yet, this reading misses the likely source of generosity, namely inequity aversion. In the risky dictator game, participants must balance two motives: the opportunity of a large gain for themselves and the resulting imbalance in payoffs between themselves and the passive player. The more they are averse to advantageous inequity, the later they should give in to the temptation of the large gain. In the prisoner's dilemma, both these motives are present as well. But they are intimately tied to a third motive. If participants cooperate, they inevitably also expose themselves to the risk of exploitation. In terms of inequity aversion, now also aversion against disadvantageous inequity comes into play. In the Fehr/Schmidt model, the α-term is generally expected to be larger than the β-term: aversion against being exploited is more pronounced than aversion against being an exploiter. Empirically for most individuals the α and the β terms are positively correlated, i.e. individuals who are strongly averse against exploiting others are usually also pronouncedly averse against being exploited themselves (Blanco et al. 2011:Fig. 1). The negative coefficient of β thus explains itself by the fact that, for most individuals, $\alpha > \beta$. It results from the fact that individuals, who are generally highly sensitive to inequity aversion, dislike being exploited even more than exploiting others. The appropriate interpretation of the negative coefficient of β is therefore not generosity, but more generally sensitivity to inequity. The more participants are sensitive to payoff differences, the more they shy away from the risk of being exploited, even if, in the prisoner's dilemma, they can only do so at the price of exposing their partner to the risk of exploitation.

6. Discussion

Since the 1960s, many researchers have experimentally tested in which way and in which magnitude manipulations of the cardinality in a prisoner's dilemma affect behaviour. Our experiment makes two contributions to this literature. It shows that cooperativeness monotonically decreases if the payoff for joint defection increases, the other payoffs held constant. Our main contribution, however, is a differentiated behavioural explanation for this effect. We

show why previous experiments have seemingly had such a hard time finding the behavioural causes. For conditional cooperators, a prisoner's dilemma is a complicated game of multiple motives. Yet if one simultaneously controls for all of them, one is able to identify this mixture.

Specifically our participants are sensitive to expected gains from cooperation. The higher these gains, the more they are likely to cooperate. As our title expresses, our participants see cooperation as a risk that may be worth taking. That this interpretation is appropriate is further corroborated by the significant positive effect of beliefs. The more they are optimistic about the fact that their partner will cooperate as well, the more they are willing to cooperate themselves. Recall that we have tested our participants on one-shot games. Therefore cooperation cannot be strategic. It must result from (conditional) cooperativeness as a personality trait. Money maximising players would have exhibited the opposite behaviour. The more they would have expected their counterpart to cooperate, the more they would have been tempted to exploit the opportunity. Actually, controlling for all remaining motivational forces, this is exactly what we find by our regressor for inequity aversion.

In a prisoner's dilemma, cooperation is dangerous. If one's counterpart defects, one not only looses gains from cooperation. One even looses the risk free payoff P, and has to live with one's worst outcome S. Even controlling for gains from cooperation and the individual degree of optimism, we find a significant effect of the individual degree of risk aversion, of loss aversion, and of their interaction. This suggests that participants do indeed sense a tradeoff: cooperation may pay; but is the risk worth taking? The more they are risk averse, and the more they are loss averse, the earlier they switch from cooperation to defection. Gains from cooperation, optimism, and risk attitudes jointly explain choices. It is also interesting that one needs measures for both risk and loss aversion plus their interaction for a significant explanation. Apparently, some participants see cooperation as an ordinary risk. Others focus on the fact that, if their counterpart defects, they even loose the safety payoff P.

Each additional variable increases the explanatory power of our regression model. More importantly even: only in the complete model, all regressors are significant. The fact that earlier attempts at explaining motives were less successful (cf. Brosig et al. 2007; Blanco et al. 2011) might have resulted from the fact that they were drawing a less complete picture.

Despite the richness of our design, it still has a number of limitations that future experiments might want to address. We had the temptation (or defection) payoff as big as the joint payoff from mutual cooperation. As the sucker payoff was zero, the society of both players did not incur an efficiency loss if one of them cooperated and the other defected. However, we do not expect this to be an important motive, given that interaction was one-shot. Accepting a payoff of zero merely on efficiency grounds seems a fairly unlikely motive. Ideally, we would have had beliefs for every level of the safety payoff P. Yet we were concerned that such a battery of tests might have overwhelmed our participants. Moreover, excessive belief elicitation has been accused to introduce a hedging motive (but see Blanco et al. 2010). Finally, while the set

of motives we have controlled for gives us significant effects for all of them, we cannot exclude that, eventually, choices in a one-shot prisoner's dilemma are affected by yet more motives. It would be particularly interesting to understand why, in our experiment, fairness preferences do not seem to have an effect. Yet this must be left to future work.

One should be cautious when extrapolating from the lab to the reality of policy problems. Yet since a host of policy problems have been modelled as prisoner dilemmas, it nonetheless makes sense to draw very tentative policy conclusions. The tragedy of the commons (Hardin 1968) is not as tragic as theory predicts (cf. Ostrom et al. 2002). Happily the more it matters since gains from cooperation are large, the more cooperation is likely. From a policy perspective, one need not necessarily change the game such that defection is no longer the best response. It may suffice to induce a sufficient fraction of addressees to take the risk of cooperation.

One aspect in particular may be open to purposeful intervention. Our results suggest that coordination on the efficient outcome does not necessarily fail because people are too greedy or too fearful. Rather, they sense that, due to individual specific levels of inequity aversion and fear, populations tend to be heterogeneous. This makes it hard for them to predict whether the risk of cooperation is worth taking. The less they are concerned by this secondary problem of prediction, i.e., the more they are optimistic about cooperativeness in the population they interact with, the more this induces them to be cooperative themselves. This has a straightforward policy implication. Interventions that reduce the uncertainty about the cooperativeness of others are likely to be effective.

Note that our data do not suggest that people expect certainty. Therefore, interventions become meaningful, even if they only substantially reduce the uncertainty about interaction partners' attitudes, short of removing the uncertainty altogether. In the field, this may be done by sorting. This might, for instance, explain why institutions frequently turn pure public goods into club goods (for background see Cornes and Sandler 1996). An alternative option is sanctions that are too weak to make it irrational for money maximisers to defect, but that are strong enough to nudge those uncertain about the type of others. This might explain why the expected value of legal sanctions is often much smaller than the expected value of violating the rule in question, and they nonetheless do not seem to be pointless.

References

AHN, K.T., ELINOR OSTROM, DAVID SCHMIDT, ROBERT SHUPP and JAMES WALKER (2001). "Cooperation in PD games. Fear, Greed, and History of Play." Public Choice **106**: 137-155.

BERTRAND, JOSEPH LOUIS FRANCOIS (1883). "Théorie mathématique de la richesse sociale par León Walras. Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot." Journal des savants **67**: 499-508.

BLANCO, MARIANA, DIRK ENGELMANN, ALEXANDER KOCH and HANS-THEO NORMANN (2010). "Belief Elicitation in Experiments. Is there a Hedging Problem?" Experimental Economics: 412-438.

BLANCO, MARIANA, DIRK ENGELMANN and HANS-THEO NORMANN (2011). "A Within-Subject Analysis of Other-Regarding Preferences." Games and Economic Behavior **72**: 321-338.

BOHNET, IRIS and RICHARD ZECKHAUSER (2004). "Trust, Risk and Betrayal." Journal of Economic Behavior & Organization **55**(4): 467-484.

BOONE, CHRISTOPHE, BERT DE BRABANDER and ARJEN VAN WITTELOOSTUIJN (1999). "The Impact of Personality on Behavior in Five Prisoner's Dilemma Games." Journal of Economic Psychology **20**(3): 343-377.

BOONE, CHRISTOPHE, CAROLYN DECLERCK and TOKO KIYONARI (2010). "Inducing Cooperative Behavior among Proselfs versus Prosocials. The Moderating Role of Incentives and Trust." Journal of Conflict Resolution **54**(5): 799-824.

BRANDTS, JORDI and ARTHUR SCHRAM (2001). "Cooperation and Noise in Public Goods Experiments. Applying the Contribution Function Approach." Journal of Public Economics **79**: 399-427.

BROSIG, JEANNETTE, THOMAS RIECHMANN and JOACHIM WEIMANN (2007). Selfish in the End?:An Investigation of Consistency and Stability of individual Behavior http://mpra.ub.uni-muenchen.de/2035/1/MPRA_paper_2035.pdf.

BRUINS, J. JAN, WIM B. LIEBRAND and HENK WILKE (1989). "About the Salience of Fear and Greed in Social Dilemmas." European Journal of Social Psychology **19**: 155-161.

COLMAN, ANDREW M. (1995). Game Theory and its Applications in the Social and Biological Sciences. Oxford England ; Boston, Mass., Butterworth-Heinemann.

CORNES, RICHARD and TODD SANDLER (1996). The Theory of Externalities, Public Goods and Club Goods. Cambridge, Cambridge University Press.

DOLBEAR, F. TRENERY and LESTER B. LAVE (1966). "Risk Orientation as a Predictor in the Prisoner's Dilemma." Journal of Conflict Resolution **10**(4): 506-515.

ENGELMANN, DIRK and MARTIN STROBEL (2004). "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." American Economic Review **94**: 857-869.

FEHR, ERNST and SIMON GÄCHTER (2000). "Cooperation and Punishment in Public Goods Experiments." American Economic Review **90**: 980-994.

FEHR, ERNST and LORENZ GOETTE (2007). ""Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment." American Economic Review **97**: 298-317.

FEHR, ERNST and KLAUS M. SCHMIDT (1999). "A Theory of Fairness, Competition, and Cooperation." Quarterly Journal of Economics **114**: 817-868.

FISCHBACHER, URS (2007). "z-Tree. Zurich Toolbox for Ready-made Economic Experiments." Experimental Economics **10**: 171-178.

FISCHBACHER, URS and SIMON GÄCHTER (2010). "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments." American Economic Review **100**: 541-556.

FISCHBACHER, URS, SIMON GÄCHTER and ERNST FEHR (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." Economics Letters **71**: 397-404.

GÄCHTER, SIMON, ERIC J. JOHNSON and ANDREAS HERRMANN (2007). Individual-Level Loss Aversion in Riskless and Risky Choices http://www.nottingham.ac.uk/economics/cedex/papers/2007-02.pdf.

GREINER, BEN (2004). An Online Recruiting System for Economic Experiments. Forschung und wissenschaftliches Rechnen 2003. Kurt Kremer und Volker Macho. Göttingen**:** 79-93.

HARDIN, GARRETT (1968). "The Tragedy of the Commons." Science **162**: 1243-1248.

HOLT, CHARLES A. and SUSAN K. LAURY (2002). "Risk Aversion and Incentive Effects." American Economic Review **92**: 1644-1655.

HONG, KESSELY and IRIS BOHNET (2007). "Status and Distrust. The Relevance of Inequality and Betrayal Aversion." Journal of Economic Psychology **28**(2): 197-213.

KAMINSKI, MAREK M. (2003). "Games Prisoners Play." Rationality and Society **15**(2): 188-217.

KERSHENBAUM, BRENDA R. and S.S. KOMORITA (1970). "Temptation to Defect in the Prisoner's Dilemma Game." Journal of Personality and Social Psychology **16**: 110-113.

KESER, CLAUDIA and FRANS VAN WINDEN (2000). "Conditional Cooperation and Voluntary Contributions to Public Goods." Scandinavian Journal of Economics **102**: 23-39.

KÖBBERLING, VERONIKA and PETER P. WAKKER (2005). "An Index of Loss Aversion." Journal of Economic Theory **122**: 119-131.

KRAMER, RODERICK M., CHARLESG. MCCLINTOCK and DAVID M. MESSICK (1986). "Social Values and Cooperative Response to a Simulated Resource Conservation Crisis." Journal of Personality **54**: 576-592.

KREPS, DAVID M., PAUL R. MILGROM, JOHN ROBERTS and ROBERT B. WILSON (1982). "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." Journal of Economic Theory **27**: 245-252.

KUWABARA, KO (2005). "Nothing to Fear but Fear Itself. Fear of Fear, Fear of Greed and Gender Effects in Two-Person Asymmetric Social Dilemmas." Social Forces **84**: 1257-1272.

LIEBRAND, WIM B. and CHARLESG. MCCLINTOCK (1988). "The Ring Measure of Social Values. A Computerized Procedure for Assessing Individual Differences in Information Processing and Social Value Orientation." European Journal of Personality **2**: 217-230.

MCCLINTOCK, CHARLESG. and WIM B. LIEBRAND (1988). "Role of Interdependence Structures, Individual Value Orientation, and Another's Strategy in Social Decision Making. A Transformational Analysis." Journal of Personality and Social Psychology **55**: 396-406.

MILINSKI, MANFRED, RALF D. SOMMERFELD, HANS-JÜRGEN KRAMBECK, FLOYD A. REED and JOCHEM MAROTZKE (2008). "The Collective-risk Social Dilemma and the Prevention of Simulated Dangerous Climate Change." Proceedings of the National Academy of Sciences **105**(7): 2291-2294.

NOVEMSKY, NATHAN and DANIEL KAHNEMAN (2005). "The Boundaries of Loss Aversion." Journal of Marketing Research **42**: 119-128.

OSTROM, ELINOR (1990). Governing the Commons. The Evolution of Institutions for Collective Action. Cambridge, New York, Cambridge University Press.

OSTROM, ELINOR, THOMAS DIETZ, NIVES DOLSAK, PAUL C. STERN, SUSAN STONICH and ELKE U. WEBER, Eds. (2002). The Drama of the Commons. Washington, National Academy Press.

POPPE, MATTHIJS and LISBETH UTENS (1986). "Effects of Greed and Fear of Being Gypped in a Social Dilemma Situation with Changing Pool Size." Journal of Economic Psychology **7**: 61-73.

RABIN, MATTHEW (2000a). "Risk Aversion and Expected-Utility Theory. A Calibration Theorem." Econometrica **68**: 1281-1293.

RABIN, MATTHEW (2000b). "Risk Aversion and Expected utility Theory. A Calibration Theorem." Econometrica **68**(5): 1281-1292.

RABIN, MATTHEW and RICHARD THALER (2001). "Anomalies - Risk Aversion." Journal of Economic Perspectives **15**: 219-232.

RAPOPORT, AMNON and DALIT ESHED-LEVY (1989). "Provision of Step-Level Public Goods. Effects of Greed and Fear of Being Gypped." Organizational Behavior and Human Decision Processes **44**: 325-344.

RAPOPORT, ANATOL (1967). "A Note on the "Index of Cooperation" for Prisoner's Dilemma." Journal of Conflict Resolution **11**: 100-103.

RAPOPORT, ANATOL and ALBERT M. CHAMMAH (1965). Prisoner's Dilemma. A Study in Conflict and Cooperation. Ann Arbor,, University of Michigan Press.

ROSENTHAL, ROBERT W. (1981). "Games of Perfect Information, Predatory Pricing and the Chain Store Paradox." Journal of Economic Theory **25**: 92-100.

ROSS, LEE and RICHARD E. NISBETT (1991). The Person and the Situation. Perspectives of Social Psychology. New York, McGraw-Hill.

SALLY, DAVID (1995). "Conversation and Cooperation in Social Dilemmas. A Meta-analysis of Experiments from 1958 to 1992." Rationality and Society **7**(1): 58-92.

SCHMIDT, DAVID, ROBERT SHUPP, JAMES WALKER, K.T. AHN and ELINOR OSTROM (2001). "Dilemma Games. Game Parameters and Matching Protocols." Journal of Economic Behavior & Organization **46**: 357-377.

SELTEN, REINHARD (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. Beiträge zur experimentellen Wirtschaftsforschung. Ernst Sauermann. Tübingen, Mohr**:** 136-168.

SELTEN, REINHARD (1978). "The Chain Store Paradox." Theory and Decision **9**: 127-159.

SIMPSON, BRENT (2003). "Sex, Fear, and Greed. A Social Dilemma Analysis of Gender and Cooperation." Social Forces **82**: 35-52.

STEELE, MATTHEW W. and JAMES T. TEDESCHI (1967). "Matrix Indices and Strategy Choices in Mixed-motive Games." Journal of Conflict Resolution **11**(2): 198-205.

STRAFFIN, PHILIP D. (1993). Game Theory and Strategy. Washington, DC, Mathematical Assoc. of America.

SUROWIECKI, JAMES (2004). The Wisdom of Crowds. Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations. New York, Doubleday :.

SWOPE, KURTIS J., JOHN CADIGAN, PAMELA M. SCHMITT and ROBERT SHUPP (2008). "Personality Preferences in Laboratory Economics Experiments." Journal of Socio-Economics **37**: 998-1009.

THALER, RICHARD H., AMOS TVERSKY, DANIEL KAHNEMAN and ALAN SCHWARTZ (1997). "The Effect of Myopia and Loss Aversion on Risk Taking. An Experimental Test." Quarterly Journal of Economics **112**(2): 647-661.

VAN LANGE, PAUL and WIM B. LIEBRAND (1990). "Causal Attribution of Choice Behavior in Three N-Person Prisoner's Dilemmas." Journal of Experimental Social Psychology **26**: 34-48.

VLAEV, IVO and NICK CHATER (2006). "Game Relativity. How Context Influences Strategic Decision Making." Journal of Experimental Psychology: Learning, Memory and Cognition **32**(1): 131-149.

WIESNER, JEROME B. and HERBERT F. YORK (1964). "National Security and the Nuclear Test-Ban." Scientific American **211**: 27-35.

Appendix

Instructions

Welcome to our experiment. Please remain quiet and do not talk to the other participants during the experiment. If you have any questions, please give us a signal. We will answer your queries individually.

Course of Events

The experiment is divided into four parts. We will distribute separate instructions for each of the four parts of the experiment. Please read these instructions carefully and make your decisions only after taking an appropriate amount of time to reflect on the situations, and after we have fully answered any questions you may have. Only when all participants have decided will we move on to the next part of the experiment. All of your decisions will be treated anonymously.

Your Payoff

At the end of the experiment, we will give you your payoff in cash. Each of you will receive the earnings resulting from the decisions you will have made in the course of the experiment. It is possible to make a loss in one part of the experiment. These losses will be subtracted from the earnings in the other parts.
Thus:
Total payment =
+ Earnings from Part 1
+ Earnings from Part 2
+ Earnings from Part 3
+ Earnings from Part 4
(min. 5€)
In Part 2, however, losses are possible, too. Should you incur losses, these will be deducted from your earnings from Part 1, Part 3, or Part 4. (The possibility of losses in Part 2 is limited, however; you will definitely receive a total payment that is on the plus side of the balance.) If you earn on the whole less than 5€, you will get a minimum payment of 5€.
We will explain the details of how your payoff is made up for each of the four parts separately. In each of the four parts, possible payoffs are given in Euro, which is the currency you will be paid in.

Part 1

The basic idea of this part of the experiment is as follows: you are anonymously paired by us with another participant. You and the other participant will make a total of eleven decisions.
Only one pair of decisions will determine your payoff. This procedure is explained below.
We will show you eleven tables that look as follows:

		Type B	
		Above	*Below*
Type A	*Above*	5€, 5€	0€, 10€
	Below	10€, 0€	z€, z€

We will let you know at the start whether you are a Type A or a Type B participant. (You will probably notice that the payments given to both types are sym-

metrical; the distinction between Type A and Type B is solely for the purpose of explaining the experiment.)
The decisions *Above* or *Below* determine the payoffs to you and the other participant. In each of the four cells of the table, the figure on the left denotes A's profit, while the figure on the right denotes B's profit.
For instance, if Type A chooses the option *Above* and Type B chooses the option *Above*, then both receive a payment of 5€. If Type A chooses *Above* and Type B chooses *Below*, then Type A receives zero profit and Type B gets 10€. The same is valid for a *Below*/*Above* constellation. Finally, if Type A chooses *Below* and Type B chooses *Below*, then both receive a payment of z€.
What does the z stand for? z is varied in the following eleven tables; all other payments remain unchanged. You have to decide on all eleven tables (*Above* or *Below*). Please mark your decision by clicking on the appropriate box shown on your screen.
You will be free to address each of the eleven tables separately, making your decisions independently of the other tables. You can also make the same decision all the time. This is entirely up to you.
Please note, once again, that only one of the eleven decision pairs will be relevant for your payoff. We will choose one of the eleven tables at random at the end. Your decision for the table that is drawn by lot and the other participant's decision for the same table determine the payoff in this part of the experiment.
Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.
Do you have any further questions?

Part 1a

This part of the experiment refers to the previous part where you made eleven decisions, "Above" or "Below". The number of participants who participated in this task will be presented to you on the screen. We ask you to estimate how many participants of the experiment selected "Above" for a particular Z (see the decision screen for detailed information). In case you make a precise estimation, you can gain 2€ in addition. If your estimation deviates by +/-2, you still gain 1€ in addition. Otherwise, you gain nothing in addition.

Part 2

The basic idea of this part of the experiment is as follows. In the following, you will be requested to make six decisions. In this part of the experiment, no other participant is paired with you. The payoffs therefore relate only to you. In each of your six decisions, you may therefore choose to play a "lottery" or decline.
What are these "lotteries" then? In these lotteries, a computer-simulated random toss of a coin determines whether you win or lose money. If the coin shows "tails" (i.e., a number), you win 6€; if it is "heads", you lose. How much you lose depends on the particular lottery. Losses vary between 2€ and 7€. If losses occur, they are subtracted from the earnings from the other parts of the experiment at the end of the experiment.
You can accept or refuse these lotteries on an individual basis, just as you can accept or refuse all. If you refuse, you will make no profit and lose nothing, i.e., your payoff will be zero. If you accept, the toss of the coin determines your payoff, as described above.
In the end, one of the six lotteries is randomly chosen, and then the payment is determined according to your decision and the coin throw for this particular lot-

tery. Thus, once again the lot decides twice in a row: first, one of the lotteries is drawn by lot, and then the toss of a coin decides whether or not you win in this lottery – on condition that you have decided to go for the lottery.
Let us first begin with some test questions. (The aim of these questions is merely to verify whether all participants have fully understood the instructions. Neither the questions nor the answers have anything to do with your final payment.) Then the screen on which your actual decisions are marked will appear.

Part 3

This part of the experiment is as follows: one Type X participant has to decide between two situations (1 or 2). His decision influences his own payoff, and the payoff of one other randomly paired Type Y participant, as follows:
Situation 1: Type X receives a payoff, determined by lot, of 5€ or 10€, Type Y receives a payoff of zero Euro. The likelihood with which Type X either receives 5€ or 10€ is systematically varied in the following table. Type X must make a decision for each of the eleven constellations (a total of 11 decisions).
Situation 2 remains the same for all 11 constellations: Type X and Type Y both receive 5€.
In this part, all participants must initially make their decisions in the role of Type X.
We will proceed with the payoff as follows:

- The lot is drawn to determine whether your payments, following your own decisions, classify you as a Type X or a (passive) Type Y. We will draw one half of the group as Type X and the other as Type Y.
- The next draw pairs each Type Y participant with a Type X participant.
- Finally, the third draw determines one single payoff-relevant situation out of the total of eleven situations. Therefore, one out of the eleven decisions emerges as the basis for payoff. With a probability of ½, it will be your own decision, and with the same likelihood it will be another participant's decision.

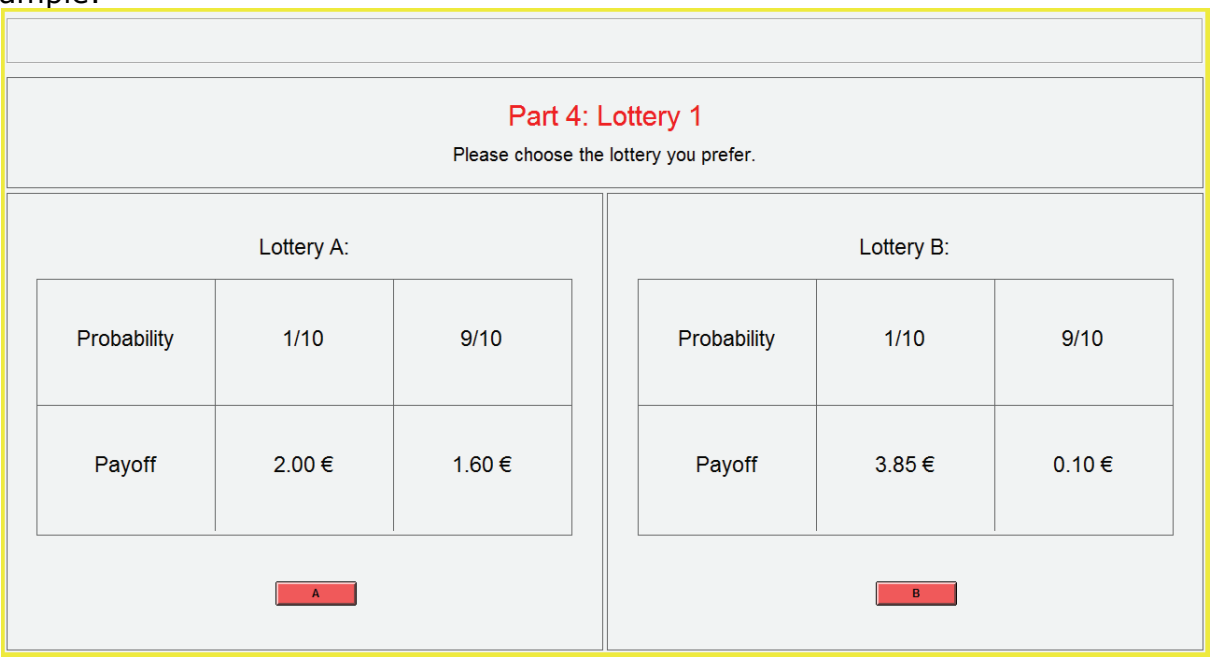
Example for Part 3

	Profit		With likelihood of
You	10€		30%
	5€		70%
Other participant	0€		100%
		1	
Your decision			
		2	
Both	5€		100%

As stated above, all participants will make eleven decisions of this kind. Please mark your decision by clicking on the appropriate box.

Part 4

In this part of the experiment, no other participant is paired with you. The payoffs therefore relate only to you. The decisions of the other participants only have an influence on their own respective payoffs.
In this part of the experiment, you are asked to decide in 10 different situations (lotteries) between option A and B. These situations will be presented to you on consecutive screens. The two lotteries each comprise 2 possible monetary payoffs, one high and one low, which will be paid to you with different probabilities. The options A and B will be presented to you on the screen as in the following example:



The computer uses a random draw program, which assigns you payments exactly according to the denoted probabilities.

For the above example, this means:
Option A obtains a payoff of 2 Euro with a probability of 10% and a payoff of 1.60 Euro with a probability of 90%.
Option B obtains a payoff of 3.85 Euro with a probability of 10% and a payoff of 0.10 Euro with a probability of 90%.
Now you have to click on the particular option you decide for.
Please note that at the end of the experiment only one of the 10 situations will eventually be paid. Yet, each of the situations can be randomly chosen with equal probability to be the payoff-relevant one.
After this, a draw will determine whether for the payoff-relevant situation the high payoff (2.00 Euro or 3.85 Euro) or the low payoff (1.60 Euro or 0.10 Euro) will be paid.