

GAINS, LOSSES, AND COOPERATION IN SOCIAL DILEMMAS AND COLLECTIVE ACTION: THE EFFECTS OF RISK PREFERENCES

WERNER RAUB* and CHRIS SNIJDERS†

*Department of Sociology, Utrecht University,
Heidelberglaan 1, 3584 CS Utrecht, The Netherlands*

We address two related issues. First, we analyze the effects of risk preferences on cooperation in social dilemmas. Second, we compare social dilemmas in which outcomes represent gains with dilemmas where outcomes represent losses. We show that predictions on gain–loss asymmetries with respect to conditions for cooperation crucially depend on assumptions concerning risk preferences. Under the assumption of risk aversion for gains as well as losses together with an assumption of decreasing absolute risk aversion, conditions for cooperation are *less* restrictive if outcomes represent losses than if outcomes represent gains. Conversely – and counterintuitively – under the assumption of S-shaped utility, conditions for cooperation are *more* restrictive if outcomes represent losses than if outcomes represent gains. We provide an experimental test of such predictions. Only a minority of subjects behaves consistent with the assumption of S-shaped utility. Furthermore, we find no empirical evidence for a general difference between cooperation in social dilemmas in which outcomes represent gains and dilemmas where outcomes represent losses. We do find evidence that risk preferences affect cooperation rates.

KEY WORDS: Collective action, conditional cooperation, gains, losses, repeated games, risk preferences, S-shaped utility, social dilemmas.

* Corresponding author.

† Useful suggestions and comments by Norman Braun, Hartmut Esser, Bernd Lahno, and Jeroen Weesie are gratefully acknowledged. Financial support was provided by the Netherlands Organization for Scientific Research (NWO) under grant PGS 50-370.

1 COOPERATION IN SOCIAL DILEMMAS AND COLLECTIVE ACTION: GAINS VERSUS LOSSES AND THE EFFECTS OF RISK PREFERENCES

Much experimental research on the Prisoner's Dilemma (PD) and other "social dilemmas" has focused on effects of variations of the payoff structure on the frequency of cooperation and related dependent variables (see Rapoport, 1974: 26–7; Colman, 1982: 118–9; Van Lange *et al.*, 1992: 14–5, for overviews and further references). Surprisingly, however, there seems to be no systematic research that compares dilemmas where outcomes represent "gains" in terms of some underlying commodity like, e.g., money with dilemmas where outcomes represent "losses". Game-theoretic and experimental research on bargaining problems has recently investigated whether bargaining over gains differs from bargaining over losses (see Camerer *et al.*, 1993). Such analyses are lacking for cooperation problems although it has been suggested that they might be useful (see Van Lange *et al.*, 1992: 14–5; Murnighan and King, 1992: 168, 181; Komorita and Carnevale, 1992: 210–1).¹ An example of a social dilemma where outcomes represent gains is the case where the status quo for a set of actors is a situation without some valuable collective good. Production of a positive amount of the collective good is a gain relative to the status quo and requires cooperation of at least some of the actors. However, each actor has an individual incentive not to contribute to the production of the collective good and to free ride on the others' contributions. An example of a dilemma where outcomes represent losses is, conversely, the case where the actors can foresee that their current situation will deteriorate through the production of a collective bad. Cooperation of all actors reduces the amount of the collective bad and, hence, reduces losses compared to the situation where

¹ We wish to compare conditions for cooperation in situations where cooperation is necessary to realize gains with conditions for cooperation in situations where cooperation is necessary to reduce losses. While this problem has been neglected in the literature, there has been some attention for another issue which should be clearly distinguished from our's, namely, "greed" versus "fear" as an instigator of defection in the PD (see Rapoport *et al.*, 1976: 104, 239; Van Lange *et al.*, 1992: 15). "Greed" refers to an actor's incentive to defect in a PD in order to unilaterally exploit a cooperative partner. Conversely, "fear" refers to an actor's incentive to defect in a PD in order to avoid exploitation through a defecting partner.

no cooperation is achieved. However, an actor minimizes his individual losses by abstaining from cooperation and free riding on the other actors' contributions for reducing the amount of the collective bad. We will refer to situations of the former type as situations where cooperation yields gains. Situations of the latter type will be referred to as situations where cooperation reduces losses (see also Section 4). More concretely, consider labor disputes and compare the situation where cooperation of employees through collective action like, e.g., a spontaneous strike would yield higher wages or better working conditions with the situation where cooperation of employees would be necessary to confine wage reductions or a deterioration of working conditions.

A systematic comparison of social dilemmas where cooperation yields gains with social dilemmas where cooperation reduces losses seems useful because it has been forcefully argued that actors tend to evaluate outcomes in terms of changes in wealth or welfare relative to some initial position rather than in terms of final positions (see, e.g., Kahneman and Tversky, 1979: 277–80, for a well-known discussion and further references). Moreover, substantial empirical evidence has been accumulated which documents that actors tend to react differently to changes of similar size depending on whether the changes represent gains or losses relative to the status quo as a reference point (see Thaler, 1992: Chapter 6 for a review). In the literature on collective action and social movements it has recently been stressed that collective action can be oriented toward gain for the participants as well as toward “the defense of existing rights or privileges that are under threat” (Walder, 1994: 2). While much theoretical work focuses on collective action for the provision of a collective good not currently enjoyed, Walder (1994) has argued from an empirical viewpoint that many historically important social movements characteristically sought to avoid or reduce loss through the production of a collective bad. According to Walder (1994: 9), historical evidence suggests an “obvious” proposition: “that loss is a more powerful motivator than gain, or that groups threatened with loss will be more likely to protest than groups that seek proactively to achieve a gain” (see Hardin, 1982: 61–5, 82–3, for an early discussion of the topic). We will test an implication of this proposition experimentally later in this paper.

Decision theorists have suggested various properties of utility functions for gains and losses. In this paper, we analyze the effects of risk

preferences on cooperation in social dilemmas. We compare three alternative assumptions on risk preference patterns for gains and losses: (i) risk neutrality for gains as well as losses, (ii) risk aversion for gains as well as losses together with an assumption of decreasing absolute risk aversion, and, finally, (iii) S-shaped utility resulting from aversion for gains and risk seeking behavior for losses. We derive implications of each of these assumptions for cooperation in repeated social dilemmas.

We show that different assumptions on risk preferences imply different and sometimes counterintuitive predictions for cooperation in repeated social dilemmas where outcomes represent gains or, respectively, losses. The basic theoretical result is that *risk aversion favors cooperation*. Not surprisingly, no gain-loss asymmetries are to be expected under the assumption of risk neutrality. We show that under the assumption of risk aversion for gains as well as losses together with an assumption of decreasing absolute risk aversion, conditions for cooperation are *less* restrictive if outcomes represent losses than if outcomes represent gains. Conversely, under the assumption of S-shaped utility, conditions for cooperation are *more* restrictive if outcomes represent losses than if outcomes represent gains. This result seems particularly counterintuitive in the light of Walder's proposition. In terms of our labor dispute example: If collective action of employees presupposes spontaneous, conditional cooperation and if utility functions are S-shaped, the prospects for collective action are better if collective action yields gains with respect to wages or working conditions than in situations in which cooperation is necessary to reduce losses with respect to wages or working conditions.

To put our analysis in perspective, we would like to mention that recent work in the social sciences based on rational choice theory frequently elaborates micro-foundations of the theory: increasingly more complex and allegedly more "realistic" assumptions on behavioral regularities are introduced (the well-known volume Hogarth and Reder (eds.), 1987, is a relatively early example for these efforts). In a sense, our analysis follows this trend in that we add to the complexity of micro-assumptions on risk preferences and, hence, properties of utility functions. We are well aware of Coleman's (1987) criticism of such tendencies. Coleman has forcefully argued that such work is frequently misguided because simple behavioral assumptions will be

adequate for most applications of the rational choice approach in the social sciences and because elaboration of micro-assumptions tends to distract from a more crucial task for social scientists, the derivation of macro-level social outcomes from individual action. We subscribe to the spirit of this argument. However, we would like to stress that our elaboration of micro-foundations *does* yield new implications for macro-level phenomena associated with cooperation in social dilemmas.

The paper is organized as follows. The next section introduces a simple model of a repeated social dilemma. We then briefly summarize basic requirements for individually rational conditional cooperation in a repeated social dilemma. The core theoretical section offers implications of different types of risk preferences for cooperation and a comparative analysis of conditions for cooperation in dilemmas where cooperation yields gains or, respectively, reduces losses. Subsequently, we provide an experimental approach towards gain–loss asymmetries in social dilemmas that allows for an empirical test of hypotheses derived from the theoretical model. A discussion concludes the paper.

2 THE REPEATED PRISONER'S DILEMMA: ASSUMPTIONS, NOTATIONS, AND DEFINITIONS

As an example of a collective action problem, we consider an infinitely or indefinitely repeated social dilemma game Γ .² For the social dilemma itself, the constituent game of Γ , we consider the standard 2-person PD.³ We assume complete information and common knowledge with

² It goes without saying that we need not claim – and do not wish to claim – that *all* collective action problems can be conceived as social dilemmas. Our analysis bears on collective action problems if a relevant subset of such problems can be reasonably approximated by situations with strategic interdependencies of the social dilemma type, a point that has indeed been frequently made in the standard literature (e.g., Taylor, 1976/1987; Hardin, 1982; and many others). Also, we do not claim that all collective action problems imply repeated instead of one-shot interactions. We believe that with respect to our example of collective action problems, a framework with repeated interactions is empirically less inappropriate than the assumption of one-shot interactions. While we consider the assumption of at least some repeated play as rather straightforward, we submit that the assumption of purely spontaneous cooperation without any external coordination and enforcement is much more problematic in many instances of collective action problems. We briefly return to this latter issue in our concluding discussion.

³ Our results can be easily generalized to, e.g., *n*-person dilemma games of the Schelling-type (Schelling, 1978: Chapter 7). See also our discussion section below.

respect to the structure of the game. Actors move simultaneously and each actor i has two pure strategies which we denote by C (“cooperation”) and D (“defection”). We characterize i ’s outcomes M as changes in wealth for i , i.e., as differences in the amount of money i owns before and after the game has been played. Hence $M = 0$ is the status quo and reference point. Outcomes depend only on the chosen strategies by the players. We use common and convenient notation for the outcomes: T for unilateral defection, R for mutual cooperation, P for mutual defection, and S for unilateral cooperation. By definition of the PD $S < P < R < T$. The PD is completely characterized by the tuple (S, P, R, T) .

Next, we turn to an actor’s utility. We assume that both actors evaluate outcomes by the same utility function U , which is likewise assumed to be common knowledge. Each actor is assumed to derive utility only from changes in own wealth. We assume that U is three times differentiable with respect to M (see note 7 below) and that wealth is desirable for each actor, hence

$$U' > 0. \quad (1)$$

Notice that (1) immediately yields

$$U(T) > U(R) > U(P) > U(S). \quad (2)$$

According to (2), D is a strictly dominant strategy for each actor so that the dilemma has a unique equilibrium such that both actors choose D . On the other hand, this equilibrium is Pareto-inefficient and both actors are better off had they both chosen C .

Now consider the repeated game Γ . This is again a game with complete information and common knowledge of the actors on the structure of the game. We assume that the PD is infinitely or indefinitely repeated in rounds $t = 1, 2, \dots$. In round t , both actors are informed on the behavior of the other actor in all previous rounds $1, \dots, t - 1$. With respect to utility functions for Γ , we assume exponential discounting of constituent game payoffs.⁴ Each actor is assumed to use

⁴ Note that this implies that we disregard possible wealth effects in the repeated game. In Section 6, we provide a new technique that allows to exclude such wealth effects in an experimental repeated game.

the same discount parameter w ($0 < w < 1$). In the following, we interpret w as a continuation probability. Hence, we interpret Γ as an indefinitely repeated PD where w is the constant probability that round $t + 1$ will be played after round t has been played, while $1 - w$ is the probability that Γ is terminated after round t has been played.

3 CONDITIONAL COOPERATION IN THE REPEATED GAME

Cooperation in the repeated game Γ is individually rational if it is supported by a subgame perfect equilibrium (spe). Standard supergame theory can be applied to derive properties of a spe of Γ such that each actor chooses C continuously on the equilibrium-path (see Fudenberg and Maskin, 1986, for a rigorous technical treatment of repeated games and Taylor, 1976/1987, as well as Axelrod, 1984, for influential applications in the social sciences). First, a spe such that both actors choose C throughout the repeated game presupposes that actors use a conditionally cooperative strategy (see, e.g., Taylor, 1987: Chapter 4). This means that an actor chooses C as long as no actor ever deviated from C in earlier rounds but, on the other hand, threatens deviations from C through own future sanctions, i.e., own subsequent D -choices.

A much stronger implication of supergame theory is (see Myerson, 1991: 327–8, for some background) that a spe such that both actors choose C throughout the repeated game exists if and only if there is a spe which consists of what are conventionally called “trigger strategies”. These strategies represent an extreme form of conditional cooperation. If actor i plays a trigger strategy, he chooses C as long as C has been chosen by both actors in all previous rounds. As soon as a deviation from C has occurred, actor i plays D in all subsequent PD's. Obviously, if both actors use a trigger strategy, a deviation from C is threatened with the most severe feasible sanction. Hence, if there is a spe such that both actors choose C throughout the repeated game, trigger strategies played by both actors are likewise a spe. A spe in trigger strategies exists if the discount parameter is large enough. The following lemma specifies the familiar requirement for a spe in trigger strategies and, hence, provides necessary and sufficient conditions for individually rational cooperation in Γ .

LEMMA 1 Γ has a spe in trigger strategies and, hence, a spe such that both actors choose C on the equilibrium-path if and only if

$$w \geq w_U^*(\Gamma) = \frac{U(T) - U(R)}{U(T) - U(P)}. \quad (3)$$

Proof See, e.g., Friedman (1986: 88–9). \square

Lemma 1 specifies a threshold $w_U^*(\Gamma)$ for the discount parameter such that trigger strategies are in spe. The equilibrium condition shows that an actor's short-run incentive to choose D , i.e., $U(T) - U(R)$, must be met by sufficient long-run costs $U(T) - U(P)$ inflicted on the deviator in each future round (weighted with their importance w). Our subsequent analysis will focus on the dependence of this threshold $w_U^*(\Gamma)$ on the utility function U . Hence, we use standard payoff-dominance arguments and assume that cooperation in the repeated game will be favored if cooperation is supported by a spe. This will allow to derive implications of risk preferences, gains, and losses on cooperation via their respective effects on $w_U^*(\Gamma)$.

4 RISK PREFERENCES AND GAIN-LOSS ASYMMETRIES

We consider three types of risk preferences and associated properties of the utility function U . Risk neutrality is associated with a linear utility function, i.e., $U'' = 0$. Risk aversion goes with a concave utility function, i.e., $U'' < 0$. Finally, risk seeking behavior implies a convex utility function, i.e., $U'' > 0$.

The following lemma provides the building blocks for our analysis. In this lemma, we compare $w_U^*(\Gamma)$ with the ratio

$$w_M^*(\Gamma) = \frac{T - R}{T - P}. \quad (4)$$

LEMMA 2 Consider a (continuous) function U . Whether $w_U^*(\Gamma)$ is larger than, equal to, or smaller than $w_M^*(\Gamma)$ depends only on the position of the point $(R, U(R))$ relative to the line through $(P, U(P))$ and $(T, U(T))$:

- $w_U^*(\Gamma) < w_M^*(\Gamma)$ if and only if the point $(R, U(R))$ lies above the line through $(P, U(P))$ and $(T, U(T))$,
- $w_U^*(\Gamma) = w_M^*(\Gamma)$ if and only if the point $(R, U(R))$ lies on the line through $(P, U(P))$ and $(T, U(T))$,

- $w_v^*(\Gamma) > w_M^*(\Gamma)$ if and only if the point $(R, U(R))$ lies below the line through $(P, U(P))$ and $(T, U(T))$.

Proof See the appendix. □

Figure 1 gives a graphical representation of Lemma 2.⁵ Our first main theoretical result follows immediately from Lemma 2. In Theorem 1, we compare the threshold $w_v^*(\Gamma)$ for the three types of utility functions associated with risk neutral, risk averse, and risk seeking preferences. Denote the threshold $w_v^*(\Gamma)$ for a linear utility function by $w_{\text{linear}}^*(\Gamma)$, for a concave utility function by $w_{\text{concave}}^*(\Gamma)$, and for a convex utility function by $w_{\text{convex}}^*(\Gamma)$. The theorem specifies the ordering of the thresholds.

THEOREM 1 (Effects of risk preferences on thresholds for conditional cooperation) *The thresholds for a spe in trigger strategies in Γ and, hence, a spe in Γ such that both actors choose C on the equilibrium path are ordered as follows for the three types of utility functions associated with different types of risk preferences:*

$$w_{\text{concave}}^*(\Gamma) < w_{\text{linear}}^*(\Gamma) = w_M^*(\Gamma) < w_{\text{convex}}^*(\Gamma). \quad (5)$$

Proof See the appendix. □

The theorem shows that the necessary and sufficient conditions for individually rational conditional cooperation in the repeated social dilemma Γ are least restrictive if both actors are risk averse and, hence, have concave utility functions, are more restrictive for risk neutral actors with linear utility functions, and are most restrictive for risk seeking actors with convex utility functions. Intuition for this result can be provided as follows. Consider the decision situation for a focal actor who assumes, according to the logic of conditional cooperation in repeated games, that the other actor uses a trigger strategy and, therefore, cooperates conditionally. By using a conditionally cooperative strategy himself, our focal actor derives utility $U(R)$ in each round of play. Alternatively, he might contemplate to defect, yielding $U(T) > U(R)$ immediately and $U(P) < U(R)$ in future rounds. Choosing to defect is profitable for the focal actor if the repeated game is played only for a few rounds, but unprofitable if the repeated

⁵The shape of the function U in Figure 1 is of course silly if U is supposed to represent a utility function, but emphasizes that Lemma 2 holds true for any function.

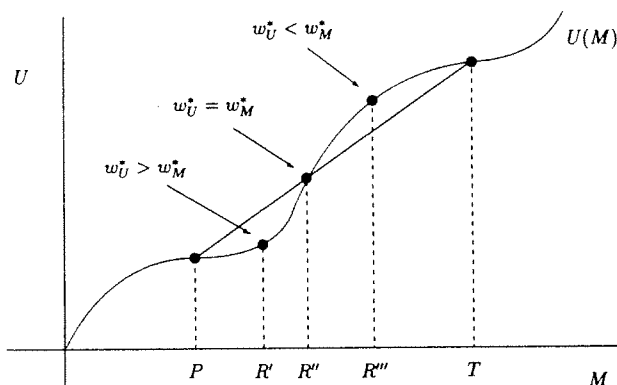


FIGURE 1 Graphical Representation of Lemma 2.

game is played for many rounds. In other words, choosing to defect can be considered as being willing to run the risk that the repeated game extends for many rounds. A risk averse actor will be least inclined to choose the gamble and defect, while an actor with risk seeking preferences will be most inclined to do so.⁶

We now compare conditions for cooperation in repeated PDs where outcomes represent gains ($M \geq 0$) and in repeated PDs where outcomes represent losses ($M \leq 0$). We consider implications of three different assumptions on utility functions for gains and, respectively, losses that correspond to three different assumptions on the actors' pattern of risk preferences in the gain- and in the loss-segment of outcomes (other assumptions are easily generated and their respective implications are likewise easily derived from Theorem 1).

Risk Neutrality for Gains and Losses

First, in the spirit of Coleman's advice to keep micro-level assumptions as simple as possible, we consider the case of *risk neutrality for gains and losses* and, hence, *linear utility functions for gains as well as losses*. In this case,

$$U'' = 0 \quad \text{for } M < 0 \quad (6)$$

⁶ Rival intuitions on the effect of risk preferences on cooperation are discussed below at the end of the section.

and

$$U'' = 0 \quad \text{for } M > 0. \quad (7)$$

Risk Aversion for Gains and Losses

Next, consider a standard economic assumption on risk preferences, namely, *risk aversion in the gain- as well as in the loss-segment* of outcomes. According to this assumption, the utility function is concave for gains as well as losses, i.e.,

$$U'' < 0 \quad \text{for } M < 0 \quad (8)$$

and

$$U'' < 0 \quad \text{for } M > 0. \quad (9)$$

A necessary and sufficient condition that allows to derive conclusions on cooperation in repeated PDs where outcomes represent gains and in repeated PDs where outcomes represent losses under the assumption of risk aversion for gains and losses is provided by a standard measure of risk-aversion (see, e.g., Arrow, 1965: 151), namely, absolute risk aversion. Absolute risk aversion is defined as

$$R_A(M) = - \frac{U''(M)}{U'(M)}. \quad (10)$$

Absolute risk aversion can be interpreted as an actor's insistence on more-than-fair odds for bets of small size. A standard assumption (Arrow, 1965: 153) is that absolute risk aversion is a decreasing function of M , i.e., $R'_A < 0$.⁷ According to this assumption, an actor's willingness to accept small bets increases with wealth, in the sense that the odds demanded diminish. We consider the implications of combining the assumption of risk aversion for gains and losses with the *assumption of decreasing absolute risk aversion*.

⁷ We have assumed that U is three times differentiable while a conventional assumption would have been that U is twice differentiable. Note that we need the assumption that U is three times differentiable only for the hypothesis of decreasing absolute risk aversion which implies that $U''' > 0$.

Risk Aversion for Gains, Risk Seeking Preferences for Losses

Finally, we address a third case of risk preferences that became popular via Kahneman and Tversky's "prospect theory" (see, e.g., Kahneman and Tversky, 1979: Section 3) and on which, e.g., Walder's proposition on loss as a more powerful motivator than gain is based. Here, risk preferences for gains differ from those for losses. Actors are assumed to exhibit *risk aversion for positive deviations from some reference point* like the status quo. Hence, it is assumed that utility is concave in the gain-segment. It is assumed that actors are *risk-seeking with respect to negative deviations from the reference point*. Hence, utility is assumed to be convex in the loss-segment. Summarizing, we have

$$U'' > 0 \quad \text{for } M < 0 \quad (11)$$

and

$$U'' < 0 \quad \text{for } M > 0. \quad (12)$$

Together, these assumptions imply *S-shaped utility functions* like the one displayed in Figure 2. Substantial empirical support seems to be available for these assumptions (see Kahneman and Tversky, 1979: 277–80, for numerous references). While S-shaped utility functions are today presumably most closely associated with prospect theory, we wish to stress that a utility function with convex as well as concave segments is *not* specific for prospect theory. Rather, such functions are likewise consistent with standard expected utility theory (see, e.g., Savage, 1954: 103–4). In fact, similar assumptions have been incorporated in earlier models based on expected utility theory. In a well-known paper, Friedman and Savage (1948) offered a related model which, however, refers to utility functions for final states rather than changes with respect to a reference point. A model that refers to changes with respect to a reference point was suggested by Markowitz (1952).

Notice further that an additional core assumption used in prospect theory – which is in itself again also consistent with expected utility theory – is the "losses loom larger than gains"-hypothesis, i.e., the assumption that the utility function for losses is steeper than the utility function for gains. Our subsequent analysis does

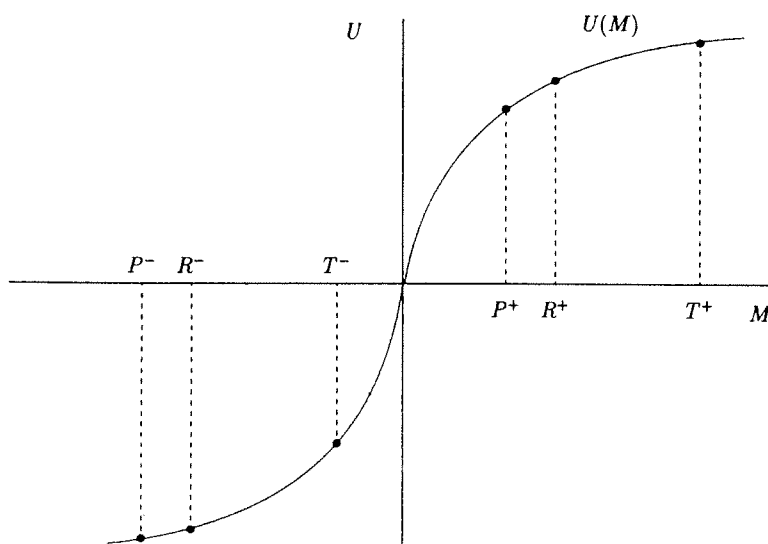


FIGURE 2 S-shaped Utility.

not depend on the adoption of an additional assumption concerning steepness. Our results hold true whether or not losses do loom larger than gains.

We now compare the conditions for cooperation in repeated PDs where outcomes represent gains and in repeated PDs where outcomes represent losses under the three different assumptions on risk preference patterns for gains and losses. We first introduce further notation that explicitly distinguishes dilemmas in the gain-segment from those in the loss-segment.

Let Γ^+ be a repeated PD characterized by the outcomes (S^+, P^+, R^+, T^+) with $P^+ \geq 0$. Hence, Γ^+ is a repeated social dilemma where cooperation yields gains. Conversely, let Γ^- be a repeated PD characterized by the outcomes (S^-, P^-, R^-, T^-) , with $T^- \leq 0$. Hence, Γ^- is a repeated social dilemma where cooperation reduces losses. Finally, define a shifted game Γ_Δ as the game which results from Γ by subtracting a constant $\Delta > 0$ from all monetary outcomes for all actors in the constituent PD of Γ . Assume furthermore that Γ^- is a shifted game of Γ^+ so that $T^- = T^+ - \Delta$, $R^- = R^+ - \Delta$, and $P^- = P^+ - \Delta$. Of course, we want to know whether the threshold $w_c^*(\Gamma^-)$ is smaller

or larger than $w_V^*(\Gamma^+)$. Note that, trivially, $w_M^*(\Gamma^+) = w_M^*(\Gamma^-)$ if Γ^- is a shifted game of Γ^+ .

Using Theorem 1, we can now derive our second main theoretical result, namely, gain–loss asymmetries under our three assumptions on risk preference patterns. The findings are summarized in the following theorem.

THEOREM 2 (Gain–loss asymmetries) *Consider a game Γ^+ where cooperation yields gains and a game Γ^- where cooperation reduces losses. Assume, moreover, that Γ^- is a shifted game of Γ^+ . In that case, the three different types of risk preference patterns generate the following orderings of the thresholds $w_V^*(\Gamma^-)$ and $w_V^*(\Gamma^+)$.*

Under risk neutrality for gains as well as for losses ((6) and (7)), the ordering is

$$w_V^*(\Gamma^+) = w_M^*(\Gamma^+) = w_M^*(\Gamma^-) = w_V^*(\Gamma^-). \quad (13)$$

Under risk aversion for gains as well as for losses combined with decreasing absolute risk aversion ((8)–(10)), the ordering is

$$w_V^*(\Gamma^-) < w_M^*(\Gamma^-) = w_M^*(\Gamma^+) < w_V^*(\Gamma^+). \quad (14)$$

Under risk aversion for gains and risk seeking preferences for losses or, respectively, S-shaped utility ((11) and (12)), the ordering is

$$w_V^*(\Gamma^+) < w_M^*(\Gamma^+) = w_M^*(\Gamma^-) < w_V^*(\Gamma^-). \quad (15)$$

Proof See the appendix. □

Remark Note that by assuming that Γ^- is a shifted version of Γ^+ we impose a condition that is stronger than necessary. For (14), we only need $w_M^*(\Gamma^-) \leq w_M^*(\Gamma^+)$, while for (15), we only need $w_M^*(\Gamma^+) \leq w_M^*(\Gamma^-)$, as can be inferred directly from the proof of Theorem 2.

The theorem establishes that whether necessary and sufficient conditions for individually rational conditional cooperation in the repeated social dilemma Γ are more (or less) restrictive if outcomes represent losses than if outcomes represent gains depends on risk preference patterns. Not surprisingly, risk neutrality for gains as well as losses implies that there are no gain–loss asymmetries. However, such asymmetries do follow from the two alternative assumptions on risk preference patterns. Under the standard economic assumption of risk aversion for gains as well as losses together with an assumption of decreasing absolute risk aversion, conditions for cooperation are *less* restrictive if outcomes represent losses than if outcomes represent gains. Conversely, under the assumption of S-shaped utility, i.e., risk aversion in the gain-segment and risk-seeking preferences in the loss-segment, conditions for cooperation are *more* restrictive if outcomes represent losses than if outcomes represent gains.

The latter result seems particularly counterintuitive in the light of Walder's proposition which, according to Walder, is based explicitly on the background assumption of S-shaped utility. Our analysis shows, contrary to Walder's proposition, that if utility is S-shaped and if collective action for the production of collective goods or the reduction of collective bads has to be based on spontaneous and self-enforcing conditional cooperation, the prospects for collective action are more favorable if the status quo and reference point is the situation without some collective good than if the status quo is the situation where some collective bad has not yet been produced. Hence, if utility is S-shaped and if collective action of employees in a labor dispute consists of a spontaneous strike where participation of at least some actors in the strike is conditional on the participation of others, then collective action for the realization of higher wages or better working conditions can be expected under less restrictive conditions than collective action for the reduction of losses with respect to wages or working conditions.

Note that labelling an actor with a convex utility function as "risk-seeking" turns out to be somewhat misleading in cases of interdependent choice. One might be tempted to think that the assumption that actors are risk-seeking would imply that actors will be inclined to cooperate more easily in a social dilemma, since they are prepared to run the risk of being "the sucker" more easily. For example, Lindenberg

(1988: 52) seems to apply such reasoning in a discussion of “the motivating power of loss”. He argues as follows:

“Following Kahneman and Tversky (1979) I assume that the gain/loss utility function is steeper for uncompensated losses than for gains. Linking this asymmetry to framing results in the following *loss hypothesis*, a hypothesis that may turn out to be of far-reaching consequences:

(a) The avoidance of uncompensated loss (i.e., the prevention of imminent loss and the reduction of recent loss) is itself a good that frames certain decision situations; (b) the likelihood that this frame dominates other possible frames in any given situation grows disproportionately with the size of the loss; (c) the costs incurred in pursuing this goal may be higher than the value of the loss itself.

For example, your wages have been cut (in your judgment) unfairly. The reduction of this loss may become a dominant goal, and you go out of your way to get even by working slower; by doing damage to factory property and by spending considerable time convincing others that some action should be taken.” (Lindenberg, 1988: 52; all emphases in the original).

Lindenberg seems to suggest that wage reductions tend to induce a “definition of the situation” by the actors concerned such that the situation without wage reductions becomes their reference point. Moreover, he seems to envision a situation where spontaneous collective action is required to reduce or avoid the collective bad. Furthermore, he seemingly suggests that in such a situation actors are particularly prone to engage in active contributions for the reduction of the collective bad (“by spending considerable time convincing others that some action should be taken”).

While such reasoning may rest on entirely plausible intuition, our analysis implies that to reach such a conclusion one actually needs to assume that actors are *risk-averse* for losses. Lindenberg suggests his reasoning is compatible with Kahneman and Tversky (1979). In fact it is not, since they typically assume actors to be *risk-seeking* for losses. Note, again, that this inconsistency does not depend on employing the assumption “that the gain/loss utility function is steeper for uncompensated loss than for gains”. In a repeated game-framework using the logic of conditional cooperation, the relevant problem for a rational actor contemplating on cooperation or defection is not whether he might become a sucker by cooperating against defecting partners. The relevant problem for such a rational actor is whether he should try a unilateral exploitation of partners who cooperate

conditionally. He has to weigh the short-term incentive for an exploitation against the expected long-term costs of such behavior. In a scenario of this type, risk aversion will favor own cooperation, while risk seeking preferences will tend to favor defection.⁸

We have mentioned that the crucial assumptions generating S-shaped utility functions are typically put forward in contexts where standard expected utility theory is confronted with alternative decision theories like prospect theory. We have stressed that our analysis is neutral with respect to controversies on standard expected utility theory and alternatives like prospect theory: S-shaped utility functions are, in principle, consistent with both alternatives. Of course, there are extensions of our analysis that *would* be inconsistent with standard expected utility theory. Assume, in particular, that essentially identical outcomes of the constituent PD with outcomes P , R , and T as defined above can be framed such that P is the reference point ("gain frame") or, respectively, T is the reference point ("loss frame"). Note that earlier discussions on framing social dilemmas as "Take-some games" and, respectively, "Give-some games" sometimes speculate on such differences and, likewise, offer some rather incoherent experimental evidence (see, e.g., Brewer and Kramer, 1986; Fleishman, 1988; Komorita and Carnevale, 1992: 210–1, as well as the more rigorous approach in Sonnemans *et al.*, 1994). Our analysis implies better chances for conditional cooperation in the repeated game under the gain frame than under the loss frame, given S-shaped utility. This implication for "social engineering", again, seems to be inconsistent with another proposition suggested by Walder (1994: 13), namely, "that ideological appeals that speak to potential adherents in terms of loss and threat shall be more effective than those that frame appeals in terms of potential gain and opportunity".

⁸ There might be another reason why Walder and Lindenberg assume a motivating power of loss. Consider S-shaped utility *and* a utility function which is steeper for losses than for gains. One might speculate that Walder and Lindenberg are simply guided by the observation that in this case utility differences between universal defection and universal cooperation or related utility differences are larger in the loss- than in the gains-segment. This intuition has been likewise put forward by Hardin (1982: 64, 82–3). However, such utility differences *per se* are neither necessary nor sufficient to generate gain-loss asymmetries in a *strategic* context. Instead *ratios* of utility differences like (3) have to be considered.

5 SHIFTING OUTCOMES WITHIN THE GAIN-SEGMENT AND WITHIN THE LOSS-SEGMENT: EFFECTS OF ABSOLUTE RISK AVERSION

Up to now, we have compared conditions for cooperation in Γ^+ and a shifted game Γ^- . We now extend the analysis and consider properties of $w_U^*(\Gamma)$ if Γ is “shifted” *within* the gain-segment or *within* the loss-segment of outcomes.

For such an analysis, we employ two variants of the threshold $w_U^*(\Gamma)$ defined in (3):

$$w_U^*(\Gamma_{\Delta}^-) = \frac{U(T^- - \Delta) - U(R^- - \Delta)}{U(T^- - \Delta) - U(P^- - \Delta)} \quad (16)$$

and

$$w_U^*(\Gamma_{\Delta}^+) = \frac{U(T^+ + \Delta) - U(R^+ + \Delta)}{U(T^+ + \Delta) - U(P^+ + \Delta)}. \quad (17)$$

Note that $w_U^*(\Gamma_{\Delta}^+)$ is the threshold for the discount parameter such that trigger strategies are in spe that is associated with a shifted game Γ_{Δ}^- . Both Γ^- as well as Γ_{Δ}^- are games where cooperation reduces losses but all losses in Γ_{Δ}^- are larger than the corresponding losses in Γ^- . Conversely, $w_U^*(\Gamma_{\Delta}^+)$ is the threshold for the discount parameter such that trigger strategies are in spe which is associated with a shifted game Γ_{Δ}^+ that results from Γ^+ by adding (instead of subtracting) a constant $\Delta > 0$ to all monetary outcomes for both actors in the constituent PD of Γ^+ . Both Γ^+ as well as Γ_{Δ}^+ are games where cooperation yields gains but all gains in Γ_{Δ}^+ are larger than the corresponding gains in Γ^+ .

We are interested in the behavior of $w_U^*(\Gamma_{\Delta}^-)$ and $w_U^*(\Gamma_{\Delta}^+)$ for increasing Δ . Increasing Δ is equivalent with shifting Γ^- or, respectively, Γ^+ within the loss-segment or within the gain-segment of outcomes, in both cases in the direction away from the origin.

Our next theorem shows that absolute risk aversion yields a necessary and sufficient condition with respect to the behavior of $w_U^*(\Gamma_{\Delta}^-)$ and $w_U^*(\Gamma_{\Delta}^+)$.

THEOREM 3 *Consider the effects of the three different types of risk preference patterns for shifting games Γ^- within the loss-segment of outcomes and shifting games Γ^+ within the gain-segment of outcomes.*

- Under risk neutrality for gains as well as losses (i.e., given (6) and (7)), $w_V^*(\Gamma_\Delta^-)$ and $w_V^*(\Gamma_\Delta^+)$ are constants for all values of T^- , R^- , and P^- as well as all values of T^+ , R^+ , and P^+ .
- Under risk aversion for gains as well as losses (i.e., given (8) and (9)) and likewise under risk aversion for gains and risk seeking preferences for losses or, respectively, S-shaped utility (i.e., given (11) and (12)), $w_V^*(\Gamma_\Delta^-)$ is a decreasing function of Δ for all values of T^- , R^- , and P^- if and only if $R'_A < 0$ and $w_V^*(\Gamma_\Delta^+)$ is an increasing function of Δ for all values of T^+ , R^+ , and P^+ if and only if $R'_A < 0$.

Proof See the appendix. □

The theorem shows that under risk neutrality the threshold for the discount parameter such that trigger strategies are in spe remains constant if the game is shifted. Under risk aversion for gains as well as losses and decreasing absolute risk aversion the threshold decreases if a repeated dilemma in which cooperation reduces losses is shifted within the loss-segment of outcomes, i.e., if a positive constant is subtracted from all (monetary) outcomes. Likewise, the threshold increases if a repeated dilemma in which cooperation yields gains is shifted within the gain-segment of outcomes, i.e., if a positive constant is added to all outcomes. The same results are obtained under S-shaped utility.

The following corollary is particularly useful for experimental tests of our results. The corollary follows immediately from Theorems 2 and 3.

COROLLARY 1 *Assume S-shaped utility together with decreasing absolute risk aversion. Consider a game Γ^+ where cooperation yields gains and a game Γ^- where cooperation reduces losses and which is a shifted game of Γ^+ . For any given parameters T^+ , R^+ , and P^+ of Γ^+ the difference $w_V^*(\Gamma^-) - w_V^*(\Gamma^+)$ decreases with increasing Δ . Likewise, if the parameters T^- , R^- , and P^- of Γ^- are assumed as given, the difference $w_V^*(\Gamma^-) - w_V^*(\Gamma^+)$ decreases with increasing Δ .*

According to this corollary, the model of conditional cooperation used in this paper together with the hypothesis of decreasing absolute risk aversion imply that differences with respect to conditions for cooperation for the production of gains versus cooperation for the reduction

of losses under S-shaped utility will be largest if, given the parameters of Γ^+ , the shift parameter Δ is chosen such that $T^- = 0$. Hence, Δ should be chosen such that $\Delta = T^+$. On the other hand, if the parameters of Γ^- are given, Δ should be chosen such that $P^+ = 0$, i.e., $\Delta = -P^-$.

Up to now, we have compared the threshold $w_V^*(\Gamma)$ for games where cooperation yields gains with the corresponding threshold for a shifted game where cooperation reduces losses. Likewise, we have studied the behavior of the threshold if games are shifted either within the gain-segment or within the loss-segment of outcomes. We did not consider games where some outcomes are gains relative to the reference point while others are losses. An analysis of these games is straightforward under the assumption of either risk neutrality for gains and losses or risk aversion for gains and losses together with decreasing absolute risk aversion. Such an analysis yields less clear results under the assumption of S-shaped utility but can be likewise based on Lemmas 1 and 2. In particular, for a given game Γ , Lemma 2 shows immediately whether the threshold $w_V^*(\Gamma)$ is smaller or larger than $w_M^*(\Gamma)$.

6 METHOD

We conducted an experiment to test two types of hypotheses, namely, those that induced work on the theoretical model we have presented in the previous sections and hypotheses that follow from this model. The experiment focused on subjects' behavior in a repeated social dilemma Γ^+ where cooperation yields gains and a shifted repeated dilemma Γ^- where cooperation reduces losses. Additionally, the experiment included checks for subjects' risk preferences over gains and losses by offering choices between appropriate certain options and risky options, i.e., gambles.

Using such a basic design, we can check, first, whether any of our three different assumptions on risk preference patterns for gains and losses is representative for a large proportion of subjects. Given the context of discovery for our model, namely, Walder's and Lindenberg's intuitive justification of their hypothesis that loss is a more powerful motivator than gain, it seems particularly interesting to see whether S-shaped utility is indeed a more or less typical pattern for utility functions.

Second, our design can be used to directly test the proposition that cooperation will be more easily achieved if cooperation reduces losses. Hence, we can test

HYPOTHESIS 1 (“loss is a more powerful motivator than gain”) *Participants will be more likely to cooperate in Γ^- than in Γ^+ .*

Of course, our model predicts that Hypothesis 1 does not hold in general but only for specific risk preference patterns. Therefore, the experiment was designed so that it allowed for a test of the crucial implication of the model itself which follows from Theorem 1 and is summarized in

HYPOTHESIS 2 (effects of risk preferences on cooperation) *Participants preferring certain options over gambles in the corresponding segments of outcomes (gains vs. losses) and, hence, exhibiting risk aversion, are more likely to cooperate in both Γ^+ and Γ^- .*⁹

Subjects

Participants were 190 male and female undergraduates. Some of them (44 out of 190) completed part of a course requirement by participating. The others reacted to an advertisement on a notice-board in the lobby of one of the buildings of Utrecht University that promised them a financial reward that could range between 6 and 36 Dutch guilders (6 Dutch guilders equals about 3.5 US dollars).

Design

We conducted eight sessions on three different days. The first two sessions (24 and 20 participants) were attended by participants that fulfilled course requirements by participating. The last six sessions

⁹ Hypothesis 2 focuses on cooperation rates in a repeated social dilemma Γ^+ where cooperation yields gains and a shifted repeated dilemma Γ^- where cooperation reduces losses. Obviously, by offering two repeated dilemmas that are shifted *within* the gain-segment or *within* the loss-segment of outcomes and using Theorem 3, we could test an analogous hypothesis, provided that we are willing to incur an assumption on decreasing absolute risk aversion.

(20, 26, 24, 24, 24, and 28 participants) were attended by participants that had reacted to the advertisement. In all sessions, the values of P^+ , R^+ and T^+ were equal to 4, 7, and 16 Dutch guilders respectively. In the first two sessions, S^+ was equal to 1, in three further sessions S^+ was equal to 0, and in the three remaining sessions S^+ was equal to 2. The payoffs S^- , P^- , R^- , and T^- in Γ^- were equal to the corresponding payoffs in Γ^+ minus 16 Dutch guilders. Note that these monetary outcomes uniquely determine the relevant discount parameter and, likewise, continuation probability for the repeated game, namely

$$w = \frac{T^+ - R^+}{T^+ - P^+} = \frac{16 - 7}{16 - 4} = \frac{0 - (-9)}{0 - (-12)} = \frac{T^- - R^-}{T^- - P^-} = 0.75.$$

All participants completed five different tasks:

Tasks 1 and 2: assessment of risk preferences. Participants' risk preferences were assessed by letting each participant make two choices. First, a choice between two options, namely (a) winning R^+ with certainty or (b) winning P^+ with probability w as defined above and winning T^+ with probability $1 - w$. Note that, e.g., subjects with a concave utility function for gains will choose the certain option (a). In the other decision situation, a subject was again offered a choice between two options, namely (c) receiving R^- with certainty or (d) receiving P^- with probability w as defined above and T^- with probability $1 - w$. Note that, e.g., subjects with a convex utility function for losses will choose the risky option (d). The order in which the choices had to be made was varied across sessions.

Task 3: comprehension check and questionnaire. Participants were quizzed about the possible consequences of different choices, and had to fill out a form asking for some personal data like age and sex.

Tasks 4 and 5: the Gains Dilemma and the Losses Dilemma. All participants played an indefinitely repeated PD Γ^+ with nonnegative payoffs S^+ , P^+ , R^+ , and T^+ (a *Gains Dilemma*), and an indefinitely repeated PD Γ^- with nonpositive payoffs S^- , P^- , R^- , and T^- (a *Losses Dilemma*). The payoffs are given above: Γ^+ was a shifted version of Γ^- . In both repeated games, the continuation probability equalled w and, hence, the termination probability $1 - w$ as indicated

above. In four sessions participants completed the two dilemmas in that order. In the other four, these two tasks were completed in the reverse order.

Note that in the design used here, each subject is given a sequence of choices, all of which have monetary consequences. This raises the problem of (controls for) wealth effects (see Davis and Holt, 1993: Chapter 8.4, for a thorough discussion). One might argue that wealth effects are indeed a serious issue in the context of our model. For example, Kahneman and Tversky (1979: 277) have suggested that outcomes are not only evaluated in terms of changes in wealth but that the utility associated with a particular change also depends on initial wealth:

"Strictly speaking, value should be treated as a function in two arguments: the asset position that serves as a reference point, and the magnitude of change (positive or negative) from that reference point. An individual's attitude to money, say, could be described by a book, where each page presents the value function for changes at a particular asset position. Clearly, the value functions described on different pages are not identical: they are likely to become more linear with increases in assets."

In terms of Kahneman and Tversky's suggestive metaphor, a subject opens up different pages of the book describing his utility function while passing through the different choices offered during the experiment. Kahneman and Tversky assume that small or even moderate variations in asset positions will have only minor effects on preferences. Nevertheless, one might wish to be able to control for such effects.

Fortunately, wealth effects can be excluded in the experiment proposed here in rather simple and efficient ways. First, the standard random-selection method (as described in Davis and Holt, 1993: 452) can be applied to exclude wealth effects due to the fact that subjects are confronted with a sequence of four decision situations (two initial choices between certain and risky prospects and two repeated PDs). According to the random-selection method, one of these decision situations is randomly selected, *ex post*, and determines the subject's actual financial earnings. Each decision situation is selected with probability $\frac{1}{4}$. Hence, choices in each situation enter a subject's overall expected utility as additively separable components. Therefore, each component should be maximized separately and potential earnings in one situation should not affect decisions in other situations.

Second, consider the somewhat more complex problem of wealth effects that arise in the course of a repeated PD. Here, one might argue that outcomes in round t affect a subject's initial position in round $t + 1$.¹⁰ To control for such wealth effects, we instructed subjects that only the *final* round of the repeated PD would be used to determine (potential) financial earnings from the repeated PD. Under such a scenario, the probability that the outcome in round t will be a subject's earning associated with the repeated PD is $w^{t-1}(1-w)$. Note that such a manipulation does not affect the strategic structure of the repeated PD. In fact (see the appendix), each term in the discounted sum of constituent game payoffs which defines expected utility for the repeated game is simply multiplied with a constant $1-w$.

A related issue concerns possible "gambling with the house money"-effects. Thaler and Johnson (1990) showed that risk preferences can be affected by prior gains and losses. The 20 Dutch guilders we provided our participants with at the beginning of the experiment could easily be considered "gambling money". Therefore, participants may have been more likely to make risk-seeking choices than they would have been if the money was theirs to begin with. We tried to minimize this effect by explicitly mentioning that the money was meant to cover their time and effort, and by letting participants put the money in their wallet immediately. Moreover, though participants may behave more risk-seekingly, it does not seem very likely that giving these 20 guilders before the experiment also affects the relation between risk preferences and cooperation.

Procedure

Per session, a maximum of 30 subjects could participate. To minimize disturbance, five experimenters were present. Participants were evenly divided over two semi-separated sides of a room. The two sides were created by a screen of about 1.5 m high across the middle of the room. Hence, participants were aware of the presence of a group of participants on the other side of the room, but could not see the participants

¹⁰ In a sense, this consideration highlights one of the problems of modelling conditional cooperation by using supergames where the same constituent game (in terms of utility) is played in each round.

in the other group during the session. Participants were instructed not to talk with each other, and they were told that the 20 Dutch guilders (about 12 US dollars) in front of them was for them, if they would take part in an experiment that could lead to a loss of a maximum of 16 Dutch guilders. After that, they were asked to read the instruction text.

The instruction text explained the different tasks and the procedure at the end of the session: All participants were going to be paid or were going to lose money according to the choices they had made in one of the tasks they had completed. The task would be determined randomly at the end of the session by picking a playing-card. If they would be payed or lose money according to their choices in either the *Gains* or the *Losses Dilemma*, they were told they would be payed or would lose money in accordance with the choices they had made *in the last round* of that particular dilemma. The instruction text furthermore explained to them that they were connected to a random participant on the other side of the room while playing the *Gains Dilemma*, to another random participant while playing the *Losses Dilemma*, and that they would not find out which participants they were connected with. The rest of the procedure of the repeated dilemmas was roughly in line with the procedure in Murnighan and Roth (1983): After all participants had made their choice for the first round of the first PD, the choices from all participants were written down by one of the experimenters and passed on to the appropriate participants on the other side of the room. After all participants had received the choices of the other player, a roulette was spun to see which pairs of participants would go on to the next round of play in the dilemma. Each participant had received a card with his or her 27 "lucky slots" on the roulette wheel. If the ball fell in one of these 27 slots, the participant would have to play another round of the dilemma (with the same partner who, of course, had the very same "lucky slots"). It was made clear in the text that if the ball would fall in the "0" slot, the wheel would be spun again. This made the probability w of continuing 0.75 (27 out of the 36 possible slots), which equals both $(T^+ - R^+)/(T^+ - P^+)$ and $(T^- - R^-)/(T^- - P^-)$. The (pairs of) participants that went on to the next round then made their choices for that following round. This procedure was repeated until no (pairs of) participants were left playing that dilemma. After the

completion of the first dilemma, participants were connected to a different participant on the other side of the room by switching the "lucky slot" cards on one side of the room and played the second dilemma in a similar way. After completion of all the tasks, participants were instructed to go individually to one of two counters in opposed corners of the room to determine the amount of money they had won or lost. The participants were not deceived in any way.

Results

All results reported here are based on the choices of the participants in round 1 of the repeated dilemmas. Obviously, the same analyses could be conducted using other dependent variables, e.g., the overall cooperation rate. None of the qualitative conclusions is affected by doing this.

Distribution of risk preferences and cooperation rates. Table 1 cross tabulates the possible patterns of risk preferences with the cooperation rates in the Gains and the Losses Dilemma.¹¹

First note that – contrary to Kahneman and Tversky's conjecture – only 23% (43/190) of all participants chose in accordance with S-shaped utility. Likewise, none of our two alternative assumptions on risk preference patterns over gains and losses seems to be representative for our subjects' revealed risk preferences. In any case, there does not seem to be much support for Hypothesis 1. We find no significant difference in cooperation rates for gains and losses based on the entire population, not even when we only consider the participants with S-shaped utility. In fact, the only pattern of risk preferences for which we do find a substantial (borderline significant) difference in favor of cooperation in the Losses Dilemma is the *seek-gains/aver-losses* pattern.¹² Note that *this* difference is inconsistent with using the

¹¹ Note that the cooperation rates in our experiment are comparable to those obtained by Murnighan and Roth (1983: 293) with a similar design.

¹² Significance tests were based on the number of subjects whose choices in the first round of the two dilemmas differed. If losses have no influence on the choice in the first round, one would expect to encounter just as many switches from cooperation to defection as vice versa. A standard binomial test can be used to test whether this is indeed the case. For the 22 subjects that were risk-seeking for gains and risk-averse for losses, we find a significant difference at $p = 0.12$ (one-tailed). All other p -values are larger.

TABLE 1
Cooperation Rates by Risk Preferences, Based on Round 1 of the Repeated Prisoner's Dilemma ($N = 190$)

	Gains Dilemma (frequencies)	Losses Dilemma (frequencies)	Percentages (frequencies)
seek-gains/seek-losses	0.29 (24)	0.23 (19)	44% (83)
aver-gains/seek-losses	0.30 (13)	0.33 (14)	23% (43)
seek-gains/aver-losses	0.14 (3)	0.36 (8)	11% (22)
aver-gains/aver-losses	0.45 (19)	0.55 (23)	22% (42)
Total	0.31 (59)	0.34 (64)	100% (190)

"aver-gains/seek-losses" is shorthand for *Risk averse for gains, risk seeking for losses*.

assumption of S-shaped utility as a basis for predicting more cooperation in the losses dilemma. Conversely, the model used here *does* predict this difference.

The effect of risk preferences and losses on cooperation. Table 2 cross tabulates risk preferences with the choice in the first round of the repeated dilemma.

Table 2 shows that the cooperation rate was larger for subjects who showed a risk-averse preference for both the Gains (38% versus 26%) and the Losses Dilemma (48% versus 26%). Based on a (one-tailed) Fisher's exact test the former difference is significant at $p = 0.054$ and the latter at $p = 0.002$.

In Table 3, we analyze the effect of risk preferences and of losses on cooperation simultaneously. As the dependent variable, we consider the choice of participants in round 1 of both the Gains Dilemma and the Losses Dilemma, treating each of these two choices as a single observation.¹³ We then construct an independent variable RISK AVERSE (REL.) that consists of participants' risk preferences on the relevant segment (gains or losses). That is, for all choices in round 1 of the Gains Dilemma the value of the variable RISK AVERSE (REL.) is a dummy-variable that describes a participant's risk preference over gains. And, likewise, for all choices in round 1 of the Losses Dilemma the value of the variable RISK AVERSE (REL.) is equal to the value

¹³ Obviously, this creates a data set with $2 \times 190 = 380$ observations. We use standard errors as introduced by Huber (1967) to account for the fact that these 380 observations are likely not to be independent. Both disregarding the dependence in the data and using a statistical model that models errors as the sum of an individual component and (independently distributed) "other noise", leads to similar results.

TABLE 2
Choices in the First Round of the Repeated Prisoner's Dilemma by
Risk Preferences ($N = 190 \times 2 = 380$)

	Defection	Cooperation	
<i>Gains Dilemma</i>			
Risk averse	53 (62%)	32 (38%)	85 (100%)
Risk seeking	78 (74%)	27 (26%)	105 (100%)
	131 (69%)	59 (31%)	190
<i>Losses Dilemma</i>			
Risk averse	33 (52%)	31 (48%)	64 (100%)
Risk seeking	93 (74%)	33 (26%)	126 (100%)
	126 (66%)	64 (34%)	190

TABLE 3
Probit Analysis on the Probability of Cooperation in the First Round
of the Repeated PD ($N = 380^{\dagger}$)

Explanatory variables	Unstandardized coefficients	(z-scores)
<i>Risk preferences</i>		
RISK AVERSE (REL.)	0.28*	(1.98)
RISK AVERSE (NOT REL.)	0.06	(0.42)
<i>Gains vs Losses</i>		
LOSS	0.10	(0.83)
<i>Controls</i>		
QUIZ	-0.00	(-0.01)
COMMON KNOWLEDGE	0.15	(0.97)
S-VALUE	0.04	(0.65)
ORDER	0.10	(0.56)
COURSE REQUIREMENT	0.69**	(3.41)
GAME THEORY	0.22	(0.19)
CONSTANT	-0.74**	(-2.61)

Log-Likelihood = -218.9; Pseudo- $R^2 = 0.08$.

*(**) = Significant at 0.05 (0.01) level, based on Huber standard errors.

[†]Two observations per subject; standard errors are computed using Huber (1967).

of the variable that describes the gamble over losses. As a consistency check, we also include a variable RISK AVERSE (NOT REL.) that consists of the risk preference that is hypothesized *not* to be relevant for the decision to cooperate in round 1. Our model predicts a positive sign of the regression coefficient for RISK AVERSE (REL.) To reexamine the effect of losses independent of the effect of risk preferences, we also include a dummy-variable LOSS that discriminates whether or not the choice considers a Gains Dilemma or a Losses Dilemma.

The data support Hypothesis 2. We do find an effect of risk aversion on cooperation in the hypothesized direction (see the variable RISK AVERSE (REL.) in Table 3): participants with risk-averse preferences are more likely to cooperate in the related social dilemma. The size of this effect can be calculated by comparing the estimated probabilities to cooperate at the mean of the dependent variable. This yields an estimated increase in the probability to cooperate of 0.10 (from 0.32 to 0.42).¹⁴ Furthermore, we do not find an effect of the variable RISK AVERSE (NOT REL.), which supports our theoretical arguments in the sense that it shows that not just any risk preference has an effect on cooperation. Again, Hypothesis 1 is not supported. Whether the dilemma is about gains or losses does not seem to have a significant effect on cooperation (see the variable LOSS in Table 3).

Miscellaneous controls. We controlled for different (possibly) intervening variables. Besides considering bivariate tests (not reported here), we incorporated several control variables in the analyses.

Participants were quizzed (in writing) about the monetary consequences of the different choices they could make. Of the 90% (170 out of 190) that answered the control questions, 78% (132 out of 170) made no mistakes at all, 18% (30 out of 170) made one mistake, and 5% (8 out of 170) made two or more mistakes. About 10% (20 out of 190) did not fill out this part of the tasks at all because they were told that given the limited amount of time the experiment would continue with the next task (the first dilemma) before they had reached these control questions. Results on cooperation rates are displayed based on the choices of all participants. Restricting the results to the 132 participants that were able to complete the control questions without mistakes in the given amount of time leads to identical results (see the variable QUIZ in Table 3).

¹⁴ An overall pseudo- R^2 of 0.08 may seem low. However, the interpretation of the pseudo- R^2 measure is not equivalent to that of the R^2 measure in standard regression analysis. In general, it is much more difficult to find pseudo- R^2 values that are large. As an illustration, consider the data set that is created by considering the percentage of cooperative moves in the first round of the repeated dilemma per session as the dependent variable. This results in a data set with 16 cases (8 sessions, each with a Gains and a Losses dilemma). We could then use the percentage of risk averse people for the relevant gamble as a predictor for this percentage of cooperative moves in a standard regression analysis (this is sloppy for several reasons, but nevertheless). This yields an (adjusted) R^2 of 0.55. The effect of risk aversion is significant at $p = 0.001$.

By assuming that the participants do not use the monetary payoff matrix to determine their choices but transform these monetary payoffs into utilities, we implicitly assume that both the subjects' own utilities and the utilities of the other subject can be considered common knowledge. As a crude check for this assumption, we asked participants after their choice in tasks 1 and 2 if they thought that most of the other participants would choose like they did. From our 190 participants, 97 thought that most of the other participants would choose as they had done in both cases. Restricting the analyses to those 97 had no significant effect on the outcomes (see the variable COMMON KNOWLEDGE in Table 3).

For all three different (sets of two) matrices we used, the monetary payoff for mutual defection plus the monetary payoff for unilateral cooperation is larger than twice the payoff for mutual cooperation. One could therefore argue that in such a Prisoner's Dilemma one should not try to get mutual cooperation started, but one had better try to get an alternating pattern of unilateral cooperation and unilateral defection started. Though this is correct in principle, the empirical relevance is immediately refuted by a look at the data. The number of alternating patterns is very small, and does not increase with an increasing payoff for unilateral cooperation. The most frequently occurring pattern seems to be an initial period of "chaos", trying to figure out what the other player is doing, followed by mutual defection after a few rounds of play. In fact, the value of S had no significant effect on the likelihood of cooperation in round 1 (see the variable S -VALUE in Table 3).

One of the tasks in the experiment was a (small) questionnaire. Participants were asked to write down their age, sex, and some other characteristics. None of these lead to any difference that approached significance (including variables AGE and MALE in Table 3's analyses leads to p -values larger than 0.69).

In the sessions where participants had to play the *Gains Dilemma* before the *Losses Dilemma*, results were not significantly different from the sessions where this order was reversed (see the variable ORDER in Table 3).

Subjects that participated to fulfill course requirements knew they were playing with other subjects from their own class and are therefore more likely to "be friendly", i.e., to cooperate in both dilemmas. We indeed find an increased probability to cooperate

(see the variable COURSE REQUIREMENT in Table 3) for these subjects.¹⁵

A final check concerned participants' familiarity with game theory. Results were not significantly different for subjects that indicated they had some knowledge of game theory (see the variable GAME THEORY in Table 3).

7 DISCUSSION

In this paper, we have offered a theoretical and experimental study of effects of risk preferences on cooperation in social dilemmas as well as a comparison of social dilemmas in which outcomes represent gains with dilemmas in which outcomes represent losses. We have assumed that outcomes in social dilemmas are evaluated in terms of changes relative to initial positions. Our theoretical findings show that predictions on gain-loss asymmetries with respect to conditions for individually rational cooperation – in the sense of conditions for the existence of subgame perfect equilibria such that each actor cooperates on the equilibrium-path – depend on assumptions concerning risk preferences for gains and losses. The fundamental theoretical result is that risk aversion favors cooperation. The requirements for individually rational conditional cooperation in a repeated social dilemma of the PD type are equivalent for dilemmas in which outcomes represent gains and for dilemmas in which outcomes represent losses if actors exhibit risk neutrality in the gain- as well as in the loss-segment of outcomes. Gain-loss asymmetries follow from other assumptions on risk preference patterns. First, if actors are risk averse in the gain- as well as in the loss-segment of outcomes and, moreover, exhibit decreasing absolute risk aversion, the requirements for cooperation in a repeated social dilemma are *less* severe if outcomes represent losses than if outcomes represent gains. Hence, assuming risk aversion for both gains and losses allows to derive a hypothesis on the “motivating power of loss”. Conversely, if actors have S-shaped utility, i.e., are

¹⁵ A separate analysis on the subjects who did not participate to fulfill course requirements gives results similar to the results displayed in Table 3: no effect of LOSS and a positive effect of risk-aversion on cooperation. A separate analysis on the data from the two sessions with subjects who *did* participate to fulfill course requirements leads to such similar results for only one of the two sessions.

risk averse in the gain-segment of outcomes but have risk seeking preferences over outcomes that represent losses, the conditions for cooperation are *more* restrictive if outcomes represent losses than if outcomes represent gains. This result seems to contradict intuitive theorizing on collective action and social movements. Even stronger, our result constitutes a serious puzzle for rational choice approaches to collective action and social movements if one is willing to make three assumptions. The first of these is that S-shaped utility is a more or less typical pattern for utility functions. The second assumption would be that Walder (1994) is correct in his interpretation of historical evidence suggesting that loss avoidance is more important than the prospect of a gain in propelling collective action. The third and final assumption would be that collective action derives from spontaneous, conditional cooperation without central coordination.

We have provided an experimental design including a sound procedure for realizing a repeated social dilemma with randomly determined termination points, checks for subjects' risk preferences, and a new procedure to exclude wealth effects from a sequence of rewards for subjects. Our data show, first, that only a minority of subjects behaves consistent with the assumption of S-shaped utility. Given our empirical evidence, it seems somewhat dubious to presuppose that S-shaped utility is a more or less typical pattern for utility functions. Second, our data do not provide empirical evidence for a general difference between cooperation in social dilemmas in which outcomes represent gains and dilemmas where outcomes represent losses. Our experimental results clearly contradict the general hypothesis that "loss is a more powerful motivator than gain." Third, we do find evidence that risk preferences affect cooperation in social dilemmas as predicted by a game-theoretic model of conditional cooperation, i.e., there is evidence that risk aversion favors cooperation. Our analysis indicates theoretical as well empirical arguments for a closer analysis of the effects of risk preferences on behavior in strategic situations. Up to now, such an analysis has been widely neglected with respect to situations of the social dilemma type (see also Ledyard, 1995: 143) as well as social interdependencies in general.¹⁶ In a future paper, we

¹⁶ Eliashberg and Winkler (1978) provide an account of the effect of exponential utility functions on the mixed equilibria of one-shot 2×2 games. See also Roth and Rothblum (1982) on the effect of risk aversion on bargaining.

hope to present results on effects of risk preferences in a broader class of social situations.

Our analysis has focused on a very simple 2-person game as a model for a social dilemma. Likewise, we have considered only one condition for individually rational cooperation in repeated social dilemmas, namely, the condition

$$w \geq w_v^*(\Gamma) = \frac{U(T) - U(R)}{U(T) - U(P)}$$

for the existence of equilibria such that everyone cooperates on the equilibrium path. With respect to possible theoretical generalizations of our analysis, we would like to mention, first, that an analysis of other "indices of cooperation" yields similar results. For example, the equilibrium condition for TIT FOR TAT-strategies in the repeated PD is

$$w \geq w_v^*(\Gamma) = \frac{U(T) - U(R)}{U(T) - U(P)} \quad \text{and} \quad w \geq \frac{U(T) - U(R)}{U(R) - U(S)}$$

Note that an analysis of the threshold $(U(T) - U(R))/(U(R) - U(S))$ produces the same results as those for the threshold $w_v^*(\Gamma)$ presented in this paper. In fact, it can be easily seen that all difference ratios except for those where numerator and denominator consist of disjunct sets of variables (e.g., $(U(R) - U(P))/(U(T) - U(S))$) lead to similar results.

Second, consider generalizations with respect to the 2-person PD which we have used as a model for a social dilemma and with respect to the repeated game. Some generalizations are straightforward. We could easily introduce heterogeneity between actors with respect to their utility functions for the constituent game and heterogeneity with respect to discount parameters. Moreover, the crucial assumptions on the strategic structure of the constituent game are that universal defection is a Pareto-inefficient equilibrium while universal cooperation would be a Pareto-improvement. Other assumptions on the constituent game, e.g., the assumption that there are only two players, that each player has only two pure strategies and the assumptions implying

that defection is a dominant strategy for each actor in the constituent game have been merely used to simplify the analysis and could be easily dropped. A far less obvious generalization of our analysis, however, would be related to the well-known folk theorems for repeated games (see Fudenberg and Maskin, 1986). This generalization would have to answer the following question: Consider the requirements for discount parameters such that a feasible and individually rational payoff vector of the constituent game (i.e., a payoff vector such that, in particular, each actor receives at least his minmax payoff) is supported by a spe of the iterated game. Are these requirements less severe if all outcomes of the constituent game are gains compared to a shifted constituent game where the outcomes are losses?

Throughout our theoretical analysis we have employed the basic assumption that individually rational cooperation in a social dilemma can only be based on spontaneous, conditional cooperation without central coordination. Of course, it is an interesting question whether an analysis would yield different predictions if one presupposes at least some degree of central coordination and, possibly, enforcement. We feel that an analysis of the possible impact of central coordination and enforcement via, e.g., political entrepreneurs or organizations is particularly relevant in the context of collective action problems and social movements. We have done some preliminary work (on which we hope to report in a future paper) in which we analyze a scenario with repeated interactions but with an additional option for each actor to incur commitments for own future cooperation in a pre-play period via, e.g., (costly) contributions to an organization that administers rewards and sanctions for cooperators and defectors throughout repeated play. In such a scenario, a *mix* of spontaneous, conditional cooperation and voluntarily incurred external coordination and enforcement can be analyzed (see also Raub and Weesie, 1993). New predictions on gain-loss asymmetries can be derived under these assumptions, e.g., one can show for such scenarios that effects of S-shaped utility for gain-loss asymmetries do depend on whether or not the "losses loom larger than gains"-hypothesis is employed: differences in steepness for different segments of the utility function matter.

Finally, consider further empirical tests and applications of the theoretical model. An obvious option, which we currently contemplate, is to rerun the experiment in Poland and with Polish subjects. This would

allow to study the robustness of our experimental results if monetary incentives relative to subjects' incomes and living costs are drastically increased. While increasing incentives might affect risk preferences (see, e.g., Kachelmeier and Shehata, 1992, for recent experimental evidence), it seems much less intuitive to us that increasing incentives also affects the relation between risk preferences and cooperation. The theory presented here focuses on this latter relation. Hence, we are somewhat reluctant with respect to the added value of such a replication.

The empirical test offered in this paper refers to the micro-level of relations between revealed risk preferences of individual actors and their cooperative or, respectively, defective behavior. Of course, such a micro-analysis could – and should – be complemented by macro-applications of the theory to, e.g., collective action and social movements. As an example, consider strikes in general and wildcat strikes in particular. More precisely, compare the situation where higher wages or better working conditions (“gains”) are the topic of labor disputes with the situation where employees are facing the prospect of wage reductions or a deterioration of working conditions (“losses”). Assumptions on risk preferences can then be used to derive predictions on the occurrence and duration of strike episodes. As another example, consider the foundation of (local) trade unions as an outcome of successful collective action in times of economic boom (“gains”) or, respectively, depression (“losses”). Again, assumptions on risk preferences allow to derive predictions on founding rates.¹⁷

REFERENCES

- Arrow, K. J. 1965. The theory of risk aversion. Reprint: pp. 141–171 in *Collected Papers of Kenneth J. Arrow*, Vol. 3, *Individual Choice under Certainty and Uncertainty*. Cambridge, MA: Belknap Press of Harvard University Press, 1984.
- Axelord, R. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Brewer, M. B. and R. M. Kramer. 1986. Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology* 50(3): 543–549.
- Camerer, C. F., E. J. Johnson, T. Ryman and S. Sen. 1993. Cognition and framing in sequential bargaining for gains and losses. In: *Frontiers of Game Theory*, edited by K. Binmore, A. Kirman and P. Tani. Cambridge, MA: MIT Press, pp. 27–47.

¹⁷ We have meanwhile identified two data sets that might allow to test such predictions. Work in this direction will be done in collaboration with Peter Hedström, Josef Brüderl, and Matthijs Kalmijn.

- Coleman, J. S. 1987. Psychological structure and social structure in economic models. In: *Rational Choice. The Contrast between Economics and Psychology*, edited by R. M. Hogarth and M. W. Reder. Chicago: Chicago University Press, pp. 181–185.
- Colman, A. 1982. *Game Theory and Experimental Games*. Oxford: Pergamon Press.
- Davis, D. D. and Ch. A. Holt. 1993. *Experimental Economics*. Princeton, NJ: Princeton University Press.
- Eliashberg, J. and R. L. Winkler. 1978. The role of attitude toward risk in strictly competitive decision-making situations. *Management Science* **24**(12): 1231–1241.
- Fleishman, J. A. 1988. The effects of decision framing and others' behavior on cooperation in a social dilemma. *Journal of Conflict Resolution* **32**: 162–180.
- Friedman, J. W. 1986. *Game Theory with Applications to Economics*. New York: Oxford University Press.
- Friedman, M. and L. J. Savage. 1948. The utility analysis of choices involving risk. *Journal of Political Economy* **56**: 279–304.
- Fudenberg, D. and E. Maskin. 1986. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* **54**: 533–554.
- Hardin, R. 1982. *Collective Action*. Baltimore, MD: Johns Hopkins University Press.
- Hogarth, R. M. and M. W. Reder (eds.). 1987. *Rational Choice. The Contrast between Economics and Psychology*. Chicago: Chicago University Press.
- Huber, P. J. 1967. The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1**: 221–233.
- Kachelmeier, S. J. and M. Shehata. 1992. Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China. *American Economic Review* **82**: 1120–1141.
- Kahneman, D. and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**: 263–291.
- Komorita, S. S. and P. Carnevale. 1992. Motivational arousal vs. decision framing in social dilemmas. In: *Social Dilemmas: Theoretical Issues and Research Findings*, edited by W. B. G. Liebrand, D. M. Messick and H. A. Wilke. Oxford: Pergamon Press, pp. 209–223.
- Ledyard, J. O. 1996. Public goods: A survey of experimental research. In: *The Handbook of Experimental Economics*, edited by J. H. Kagel and A. E. Roth. Princeton, NJ: Princeton University Press, pp. 111–194.
- Lindenberg, S. 1988. Contractual relations and weak solidarity: The behavioral basis of restraints on gain-maximization. *Journal of Institutional and Theoretical Economics* **144**: 39–58.
- Markowitz, H. 1952. The utility of wealth. *Journal of Political Economy* **60**: 151–158.
- Murnighan, J. K. and T. R. King. 1992. The effects of leverage and payoffs on cooperative behavior in asymmetric dilemmas. In: *Social Dilemmas: Theoretical Issues and Research Findings*, edited by W. B. G. Liebrand, D. M. Messick and H. A. Wilke. Oxford: Pergamon Press, pp. 163–182.
- Murnighan, J. K. and A. E. Roth. 1983. Expecting continued play in Prisoner's Dilemma Games. *Journal of Conflict Resolution* **27**: 279–300.
- Myerson, R. B. 1991. *Game Theory. Analysis of Conflict*. Cambridge, MA: Harvard University Press.
- Rapoport, A. 1974. Prisoner's Dilemma – Recollections and observations. In: *Game Theory as a Theory of Conflict Resolution*, edited by A. Rapoport. Dordrecht: Reidel, pp. 17–34.

- Rapoport, A., M. J. Guyer and D. C. Gordon. 1976. *The 2 × 2 Game*. Ann Arbor, Mich.: University of Michigan Press.
- Raub, W. and J. Weesie. 1993. Symbiotic arrangements: A sociological perspective. *Journal of Institutional and Theoretical Economics* **149**: 716–724.
- Roth, A. E. and U. G. Rothblum. 1982. Risk aversion and Nash's solution for bargaining games with risky outcomes. *Econometrica* **50**(3): 639–647.
- Savage, L. J. 1954. *The Foundations of Statistics*. 2nd. and rev. ed., New York: Dover 1972.
- Schelling, T. C. 1978. *Micromotives and Macrobehavior*. New York: Norton.
- Sonnemans, J., A. Schram and T. Offerman. 1994. Public Good Provision and Public Bad Prevention: The Effect of Framing. *Mimeo*, University of Amsterdam.
- Taylor, M. 1987. *The Possibility of Cooperation*. Cambridge: Cambridge University Press. Rev. ed. of *Anarchy and Cooperation*. London: Wiley 1976.
- Thaler, R. H. 1992. *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. New York: Free Press.
- Thaler, R. H. and E. J. Johnson. 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choices. *Management Science* **36**: 643–660.
- Van Lange, P. A. M., W. B. G. Liebrand, D. M. Messick and H. A. Wilke. 1992. Introduction and literature review. In: *Social Dilemmas: Theoretical Issues and Research Findings*, edited by W. B. G. Liebrand, D. M. Messick and H. A. Wilke. Oxford: Pergamon Press, pp. 3–28.
- Walder, A. G. 1994. Implications of Loss Avoidance for Theories of Social Movements. *Mimeo*, Harvard University.

APPENDIX

Proof of Lemma 2 and Theorem 1 Consider the line l through $(P, U(P))$ and $(T, U(T))$ in Figure 1. It is defined by

$$l: y = \frac{U(T) - U(P)}{T - P} (x - P) + U(P).$$

If U is concave, $(R, U(R))$ lies above this line, and we have

$$U(R) > \frac{U(T) - U(P)}{T - P} (R - P) + U(P),$$

which we can rearrange to

$$\frac{T - R}{T - P} > \frac{U(T) - U(R)}{U(T) - U(P)}.$$

Similarly, if U is convex, we have

$$\frac{T-R}{T-P} < \frac{U(T)-U(R)}{U(T)-U(P)}.$$

Finally, note that

$$\frac{T-R}{T-P} = \frac{U(T)-U(R)}{U(T)-U(P)}$$

if $(R, U(R))$ lies on the line through $(P, U(P))$ and $(T, U(T))$. This completes the proof of Lemma 2. Combining these inequalities leads us to the ordering of thresholds as in (5) which completes the proof of Theorem 1. \square

Proof of Theorem 2 The orderings of thresholds in (13) and (15) follow directly from Theorem 1. For (14), consider a concave utility function and the set Γ_Δ of shifted games that result from a given game Γ by adding a constant $\Delta > 0$ to all monetary outcomes for both actors in the constituent PD of Γ . Note that Γ^+ is then a shifted game of Γ^- for appropriate Δ . We first prove the following lemma.

LEMMA 3 U' is convex in $U(x) \Leftrightarrow R'_A < 0$.

Proof Let $z = U(x)$. Then $U'(x) = U' \circ U^{-1}(z)$. By straightforward computation we get that

$$\frac{d}{dz} U' \circ U^{-1}(z) = -R'_A(x),$$

$$\frac{d^2}{dz^2} U' \circ U^{-1}(z) = -\frac{R'_A(x)}{U'(x)}.$$

Hence, $R'_A < 0$ is equivalent to $d/dz U' \circ U^{-1} < 0$ and $d^2/dz^2 U' \circ U^{-1} > 0$. The latter is equivalent to $U' \circ U^{-1}$ being strictly convex in z , or, since U is a bijection, U' being strictly convex in $U(x)$, which completes the proof of Lemma 3.

Note that $dw_v^*/d\Delta > 0$ is equivalent to

$$\frac{U(T+\Delta) - U(R+\Delta)}{U(T+\Delta) - U(P+\Delta)} > \frac{U'(T+\Delta) - U'(R+\Delta)}{U'(T+\Delta) - U'(P+\Delta)}.$$

First, suppose that $R'_A(x) < 0$ for all $x > 0$. By Lemma 3, U' is strictly convex in $U(x)$. But then we have that

$$U'(R+\Delta) < \frac{U'(T+\Delta) - U'(P+\Delta)}{U(T+\Delta) - U(P+\Delta)} \\ \{U(R+\Delta) - U(P+\Delta)\} + U'(P+\Delta),$$

for all P, R, T , and Δ . Rearranging the above leads to

$$\frac{U(T+\Delta) - U(R+\Delta)}{U(T+\Delta) - U(P+\Delta)} > \frac{U'(T+\Delta) - U'(R+\Delta)}{U'(T+\Delta) - U'(P+\Delta)}$$

and hence we proved that $R'_A < 0$ implies $dw_v^*/d\Delta > 0$ for all P, R , and T .

Second, suppose that there exists an x such that $R'_A(x) \geq 0$. By Lemma 3, U' is *not* strictly convex in $U(x)$. This implies that there exist P, R, T , and Δ such that

$$U'(R+\Delta) \geq \frac{U'(T+\Delta) - U'(P+\Delta)}{U(T+\Delta) - U(P+\Delta)} \\ \{U(R+\Delta) - U(P+\Delta)\} + U'(P+\Delta),$$

which we can rearrange to

$$\frac{U(T+\Delta) - U(R+\Delta)}{U(T+\Delta) - U(P+\Delta)} \leq \frac{U'(T+\Delta) - U'(R+\Delta)}{U'(T+\Delta) - U'(P+\Delta)}.$$

Hence, there exist P, R, T , and Δ such that $dw_v^*/d\Delta \leq 0$, which completes the derivation of (14) and, hence, the proof of Theorem 2. \square

Proof of Theorem 3 The proof of the theorem for risk neutrality and for risk aversion for gains as well as losses follows directly from

Theorem 2. The proof of the theorem for S-shaped utility proceeds similar to the derivation of (14) in the proof of Theorem 2. In fact, the derivation of (14) directly yields the proof for S-shaped utility and the set of games Γ_{Δ}^+ by adding upper indices and writing “ T^+ ” for “ T ”, etc. The proof for S-shaped utility and the set of games Γ_{Δ}^- is analogous. \square

Control for wealth effects in the iterated PD If actors receive earnings from each round of play and if there are no wealth effects, actor i 's expected utility for Γ , given some vector σ of supergame strategies, is $EU(\sigma) = \sum_{t=1}^{\infty} w^{t-1} EU_t$, where EU_t is i 's expected utility in round t , given σ and given that round t is played. On the other hand, if actors receive only the earning from the final round of play, we have

$$EU(\sigma) = \sum_{t=1}^{\infty} w^{t-1} (1-w) EU_t = (1-w) \sum_{t=1}^{\infty} w^{t-1} EU_t.$$

Hence, the incentive structure of the game is unchanged while potential earnings in round t cannot affect decisions in round $t+1$. \square

ISCORE Papers

The ISCORE Papers contain papers, lectures, pre-publications, and reprints by members or visitors of ISCORE. Codebooks of data collections and descriptions of computer software developed by members of the ISCORE group are also included. We maintain a mailing list of those interested in ISCORE's activities, including its seminars and papers. If you want to receive a complete list of the papers, receive one or more of the papers (free of charge) or be included on the mailing list, please send a request to: ISCORE, Department of Sociology, Postbox 80.140, 3508 TC Utrecht, The Netherlands. Email: W.Raub@fss.uu.nl
<http://www.fss.uu.nl/soc/iscore/>