CORVINUS JOURNAL OF SOCIOLOGY AND SOCIAL POLICY VoL.7 (2016) 2, 55-76. DOI: 10.14267/CJSSP.2016.02.03

SOCIAL DILEMMAS AS EXCHANGE DILEMMAS

JACOB DIJKSTRA[1] – MARCEL A.L.M. VAN ASSEN

ABSTRACT: *We develop a new paradigm to study social dilemmas, called exchange dilemmas. Exchange dilemmas arise from externalities of exchanges with third parties, and many real-life social dilemmas are more accurately modeled as exchange dilemmas rather than prisoner's dilemmas. Building on focusing and framing research, we predict that defection is omnipresent in exchange dilemmas, which is corroborated in two very different experiments. Our results suggest that the fundamental problem of cooperation in many real-life social dilemmas may be more severe and harder to solve than suggested by traditional Prisoner's Dilemma research, due to the presence of third parties. Directions for future research are suggested, focusing on relations with third parties.*

KEYWORDS: *cooperation, social dilemmas, social exchange, externalities, experiments, prisoner's dilemma*

Explaining cooperation in social dilemmas among unrelated individuals is a major problem in the behavioral sciences (Buchan et al. 2002; Dawes, 1980; Fehr et al. 2002, 2003; Kollock, 1998; Willer, 2009). The Prisoner's Dilemma Game (PDG) is arguably the most frequently used theoretical paradigm for studying this problem (Axelrod, 1984; Kanazawa, 2007; Macy, 1991; Nowak et al. 1993, 2005; Poundstone, 1992; Takahashi, 2000). Other well-known paradigms used for studying cooperation are the Ultimatum Game (e.g., Thaler 1988), the Trust Game (e.g., Snijders and Keren 2001), and the Public Good Game (e.g. Fehr and Gächter 2000, 2002). These models share with the PDG the important property that in finite interactions 'free riding' (i.e. not cooperating with one's partner(s)) is an individual's most rewarding strategy, irrespective of the behavior of others. Since reciprocal cooperation yields an outcome that is Pareto superior to the

1 Jacob Dijkstra is assistant professor at University of Groningen / ICS, Marcel van Assen is professor at Utrecht University/ICS and assistant professor at Tilburg University; e-mail: j.dijkstra@rug.nl

outcome resulting from a total lack of cooperation, these paradigms are said to embody *social dilemmas*.

In the present paper we employ a new paradigm to study social dilemmas. We examine social dilemmas arising from externalities of exchanges with third parties, and call them *exchange dilemmas*. We contend that many real-life social dilemmas can be modeled as exchange dilemmas, and we provide well-known examples of exchange dilemmas. We demonstrate in this paper that the exchange dilemma is a versatile and fruitful paradigm for studying social dilemmas and can enhance our understanding of real-life social dilemmas. We argue that defection rather than cooperation is omnipresent in exchange dilemmas, a prediction that is corroborated in two experiments.

Exchange dilemmas

An exchange is said to have externalities if it has direct consequences (either positive or negative) for the well-being of actors not themselves partners to the exchange. An instructive example of such exchanges with externalities is found in the context of collective decision making. Two political parties may agree to exchange their voting positions concerning two issues that have to be decided upon in Parliament (e.g., Dijkstra et al. 2008; Thomson et al., 2006; Udehn, 1996). Since this "logrolling" changes the eventual outcome of the votes on both issues, the exchange directly affects the well-being of other political parties not involved in the agreement.

Exchanges with externalities can give rise to social dilemmas (Dijkstra, 2009). The intuition is that if exchanges in a social system (such as a group or a social network) have large negative externalities for others, potential benefits from the exchange may be nullified by losses due to externalities. Those losses may even outweigh the benefits and make all actors *worse off* after all exchanges have been carried out, compared to a situation in which no exchanges had taken place. However, since for each actor every possible exchange in which he is a partner is individually profitable (while the negative externalities caused by it are borne by others), all actors have an incentive to complete as many social exchanges as they can. Thus, the actors are involved in a social dilemma: a situation in which individually rational behavior leads to collectively undesirable outcomes (Dawes, 1980). Van Assen et al. (2003) and Dijkstra et al. (2008) discuss this dilemma situation in the context of collective decision making.

Characteristic of social dilemmas arising from exchanges with externalities, or *exchange dilemmas*, is that they arise from profitable exchange opportunities with third parties. A classic example is the in-shore fisheries in Bodrum and in

the Bay of Izmir, Turkey, discussed by Ostrom (1990). To prevent overfishing local fishers have to cooperate (i.e., restrict their catch) in a situation resembling a (*N*-person) PDG (Hardin, 1968). Fishermen catch their fish in order to exchange them with retailers and consumers, and these exchanges can thus be said to have negative externalities for other fishermen. The third parties in this particular exchange dilemma, the consumers and retailers of fish, have an impact on the degree of cooperation that the fishermen reach. Moreover, they have a stake in it as well, as cooperation between the fishermen means a shortage of fish at the market. Hence, when the fishermen cooperate, third parties can be expected to raise the prices they are willing to pay for fish, thereby trying to entice the fishermen to defect and catch more fish. In fact, Ostrom observes that fishing communities with easy market access (as is the case with the Bay of Izmir fishermen) are more exposed to such temptation by third parties than fishermen without (the Bodrum fishermen). Accordingly, the fishing communities with easy market access have a harder time sustaining cooperation.

Consider another well-known example of exchange dilemmas: cartels, such as OPEC. Cooperation between the oil-producing countries generally entails some kind of restriction on voluntary output. Successful cooperation results in the increased scarcity of oil and higher prices, which in turn leads to high profits for oil suppliers. Given that other countries also restrict their output (thus increasing the price of oil), each individual country is tempted to produce and sell more oil than agreed upon. The social dilemma in this case is clearly affected by the interactions between the oil-producing countries and the buyers of oil, who are the third parties. These interactions can be seen as exchanges with negative externalities for other OPEC countries. Third parties benefit from a low oil price and thus want suppliers to defect. Moreover, they are capable of affecting supplier profits through the prices they offer in their exchanges, thereby increasing the temptation to defect from the agreement for each OPEC country. Therefore, any sustainable cooperation between oil suppliers must be able to overcome the effects of the tempting behavior of third parties.

We propose that real-life social dilemmas can be modeled as exchange dilemmas whenever they involve third parties, the social exchanges with whom create the dilemma via the externalities associated with them, and who have both the incentive and the opportunity to affect the payoffs of the actors in the dilemma.

The PDG and exchange dilemma as model of social dilemmas

One standard model of social dilemmas is the PDG. It is instructive to compare the PDG and exchange dilemma representations of social dilemmas

using the original anecdote by Albert W. Tucker. In Tucker's anecdote, two men are charged with a joint violation of law and are held separately by the police. The police tell each of them that if one confesses while the other does not, the former will be set free and the latter will be severely punished. If both confess, both will receive moderate punishment. Finally, if neither confesses, the police have enough evidence to send them both to prison on a lesser charge, resulting in mild punishment. The dilemma consists of the fact that it is individually rational for both men to confess, since given the choice of the other prisoner an individual always reduces the severity of his own sentence by confessing.

The resultant dilemma for the prisoners is typically represented as a 2×2 game between the prisoners, such as is illustrated in Table 1.

Table 1. *The Prisoner's Dilemma Game*

		B	
		Cooperate	Defect
A	Cooperate	1,1	3,0
	Defect	0,3	2,2

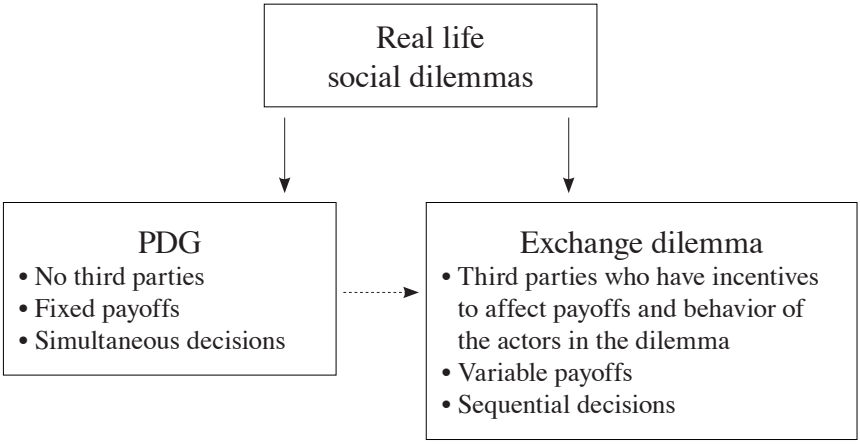
Note: Prisoners A and B can each choose to Cooperate (not confess) or to Defect (confess). Each cell gives the years in prison for A, and the years in prison for B (source: Poundstone, 1992: 118).

In Table 1 both prisoners can either cooperate (not confess) or defect (confess), resulting in 2 years' imprisonment for both if they both defect (confess to the police; lower right cell) and 1 year of imprisonment for both if they both cooperate (refuse to confess to the police; upper left cell). Unilateral defection results in 0 years of imprisonment for the confessor and 3 years for the prisoner who does not confess. Defection (i.e., confession to the police) is the prisoner's best choice since it reduces the sentence by one year, but if both defect both end up in prison for 2 years instead of

Whereas the original anecdote is often transformed into the 2×2 PDG in Table 1, the anecdote is naturally transformed into an exchange dilemma. Contrary to the case in Table 1, in the anecdote there is an *active* and *interested* third party: the police. The police structure the situation (e.g., putting the prisoners in separate cells), and offer the prisoners a bargain (in terms of prison years for cooperation and defection) *with the goal of having the prisoners confess*. And it is the interaction of the prisoners with the police that constitutes the cooperation problem for the prisoners. These interactions can be conceptualized as exchanges between the police and the prisoners, in which the former can offer fewer years in prison in return for a confession by the latter. Since such a confession entails the

conviction of the other prisoner, however, this is an exchange between prisoner and police (third party) with negative externalities. In these exchanges not only do the incentives for the prisoners matter, as the standard 2×2 PDG would suggest, but also the incentives for the police. No police officer will simply stop the interrogation when the prisoners refuse to confess: she will make alternative offers to them (by varying the length of the prison sentence), tempting them to confess. Note how this implies that the payoffs for the prisoners are variable rather than fixed.

Figure 1. *The social dilemma and the PDG and exchange dilemma*



The original social dilemma and both the PDG and exchange dilemma are schematically presented in Figure 1 along with some of their characteristics. The solid arrows signify that both the PDG and exchange dilemma are models of the social dilemma. In our experience, many researchers forget that the PDG is merely *a model* of social dilemmas, arguing that the exchange dilemma is *derived from* a PDG (cf. the dashed arrow in Figure 1). Typical comments we receive are "*you should directly test the effects of including an interested third party in the PDG*", "*how does converting the traditional binary (cooperate/defect) PD paradigm into a continuous one affect the results?*", and "*because of the third parties the situation is not seen as a PDG by the actors*". We stress, however, that the only right of existence of the PDG is that it is a *model* of the social dilemma. The exchange dilemma is another, and as Figure 1 shows, a different model of the social dilemma. Both models have their merits, and should be evaluated in comparison with the social dilemmas they model.

Overview

We first explain how exchange dilemmas can be modeled. Then we derive predictions concerning actor behavior and cooperation in exchange dilemmas. Our main prediction is that defection in the exchange dilemma is omnipresent, because actors in the dilemma are primarily focused on creating profitable exchanges with third parties rather than on the negative externalities resulting in the dilemma. Our prediction is verified in two experimental studies, Study 1 and Study 2. The two major differences between the studies are how they were run (face-to-face, paper-and-pencil in Study 1 versus web-based with participants in separate cubicles in Study 2) and the interaction structure (one third party in Study 1 versus two third parties in Study 2). Finding similar results from both studies would increase our confidence in the effectiveness of our experimental manipulation (i.e., including social exchange relations with third parties). In addition, similar results from both studies would also enhance our confidence in the working of the theoretical mechanisms in real-life exchange dilemmas, and enhance the external validity of our conclusions. The results of both studies corroborate the main prediction that defection is omnipresent in real-life social dilemmas represented as exchange dilemmas. The implications of our findings are discussed in the final section.

MODELING EXCHANGE DILEMMAS

Following a long tradition of experimental research in social exchange in sociology and social-psychology (e.g., Cook et al., 1983; Molm, 1997; Willer, 1999), we model exchange relations using mutually profitable bilateral exchanges of resources. In explaining how introducing externalities in a social exchange can induce a social dilemma situation, we describe the exchange dilemma used in Study 1.

In Study 1 both actors A and C have the opportunity to change with B; a situation represented by the so-called 3-Line exchange network of Figure 2a, in which the lines represent the possibilities (but not obligations) for exchange between connected actors. Actor B can change with either A or C, or both. The actors are endowed with different amounts of different resources. This is depicted in the third row of Table 2. There it is shown that A and B have 60 and 10 units of resource X, respectively, whereas B and C have 30 and 60 units of resource Y, respectively. Actors' payoffs for a unit of X and Y are presented in the fourth row of Table 2. A, B, and C obtain 1, 3, 5 money units, respectively, for 1 unit of X, whereas all actors obtain 1 money unit for 1 unit of Y. Thus, A and B can make

a mutually profitable exchange by trading B's 30 units of resource Y for 10 to 30 units of A's resource X. Similarly, B and C can complete a mutually profitably exchange by trading B's 10 units of resource X for 30 to 50 units C's resource Y.

Figure 2a. *The 3-Line network of Study 1*

Figure 2b. *The box network of Study 2*

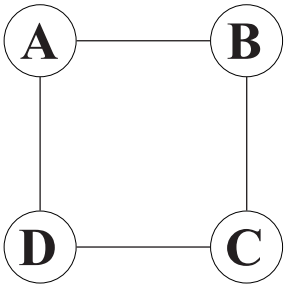


Table 2. *Endowments and payoffs of participants in Experiment 1 and Experiment 2*

Participants →		A		B		C		D	
Resources →		X	Y	X	Y	X	Y	X	Y
Study 1	Endowments	60	0	10	30	0	60		
	Payoffs	1	1	3	1	5	1		
Study 2	Endowments	1	0	0	48	1	0	0	48
	Payoffs	24	1	48	1	24	1	48	1

The exchange dilemma is created by introducing *negative externalities* of exchanges between A and C. By design of Study 1, whenever actor A transfers 1 unit of X to B, actor C loses 5 points (C's payoff for one unit of X). Conversely, whenever actor C transfers 1 unit of Y to B, actor A loses 1 point (A's payoff for one unit of Y). Thus, it is *as if the A and C actors are drawing the resources they use in exchanges with B from a common pool*, like the fishermen investigated by Ostrom (1990). The resources they receive from B, however, are private, for instance representing the money obtained by a fisherman (A or C) from consumers (B). Similarly, the exchange dilemma in Study 1 can be seen as an

example of a collective decision situation where (i) A and C disagree about two issues *X* and *Y*, (ii) *B* moves his position on X to A and on Y to C, resulting in a loss for C and A, respectively, (iii) there exists a loss that is not compensated by the private gain they obtain from their exchange with B.

The negative externalities imply the payoffs for A and C as shown in Table 3, as a function of *x* and *y* transferred by A and C, respectively. If A transfers *x* to B in exchange for B's 30 units of Y, A obtains 30 – *x*, whereas C loses 5*x* (lower left cell). If C transfers *y* to B in exchange for B's 50 units of Y, C obtains 50 – *y*, whereas A loses *y* (upper right cell). The payoffs resulting when both exchanges are completed are obtained by adding the payoffs of the two separate exchanges (lower right cell).

The exchange dilemma in Table 3 corresponds to a social dilemma caused by the exchange opportunities with third party B:

(i) completing an exchange always increases an actor's payoffs, since exchange is mutually profitable (i.e., for both the actor and the third party), and
(ii) A's loss, created by an exchange of C with B, cannot be compensated by A exchanging with B, and similarly, for C.

Table 3. *The Exchange Dilemma induced by Condition 2 of Study 1*

		C	
		Not Exchange	Exchange
A	Not Exchange	0, 0	–y, 50–y
	Exchange	30–x, –5x	30–x–y, 50–5x–y

Note: x and y are the number of units of X and Y transferred to B, respectively; $10 \leq x \leq 30$, and $30 \leq y \leq 50$, for mutually profitable exchanges.

Consequently, the order of payoffs in the four cells corresponds to that of a social dilemma, with defection (exchanging with B) being the best alternative for both A and C, and mutual cooperation (both not exchanging) yielding better payoffs to both A and C than when both defect. Moreover, contrary to the traditional 2×2 PDG of Table 1, the social exchange relationships with the third party are explicitly modeled (see again Figure 1): a third party (B) is present, the interactions with whom constitute the choices (exchange or not) for the actors in the dilemma (A and C), and who (B) has the incentives and the possibilities to influence the payoffs of the actors A and C by offering more or less profitable exchanges to them.

THEORY AND MAIN HYPOTHESIS

Experimental studies using the PDG conclude that cooperation in social dilemmas is quite common. Sally (1995) found an average cooperation rate of almost 50%, reviewing 130 treatments with different experimental conditions. Factors affecting cooperation include the possibility of pre-play communication (increasing cooperation by about 40%), the size of the 'temptation payoff' (i.e., the payoff earned by a unilateral defector; the smaller the size, the greater the cooperation rate), repeated play (the cooperation rate decreases as more and more rounds are played), and whether or not payoffs are monetary (monetary payoffs elicit on average about 20-25% more cooperation than non-monetary ones).

Notwithstanding the sizable cooperation rate found in PDG research, our main prediction is that there will hardly be any cooperation in exchange dilemmas. Two reasons underlie this expectation. First, framing and focusing render cooperation between the actors in an exchange dilemma unlikely. Evidence about framing (Kahneman et al. 1979, 1981; Levin et al. 1998) and focusing (Legrenzi et al. 1993) shows that differences in the way choices are represented to and conceived by individuals have large effects on these choices.

From the perspective of exchange dilemmas, a particularly useful distinction is the one between 'valence framing' and 'label framing' (e.g., Dufwenberg et al. 2011). Valence framing concerns the fact that different presentations can put the same information in either a 'positive' or a 'negative' light. In valence framing, the participant's reference point is shifted, for example by presenting the same information alternatively as a gain or as a loss. Label framing concerns 'mere wording', leaving the participant's reference point unchanged. For instance, Liberman et al. (2004) show how the name attached to a PDG (either 'Community Game' or 'Wall Street Game') can strongly affect participants' choices.

According to cognitive psychological theory, individuals construct a *mental model* of the social dilemma at hand, containing salient aspects of the situation, but not necessarily all relevant information from an incentive point of view (Johnson-Laird 1983). Focusing implies that different presentations of the same information lead to *different* mental models, which in turn determine how the social dilemma is framed by the individual: which norms of appropriate behavior are invoked (if any), which consequences are perceived of as costs, and which as gains, etc.? Different frames may then lead to different decisions and behavior.

Applied to the exchange dilemma this reasoning implies that each actor ('prisoner') is focused on making a potentially profitable deal with the third party ('police'), rather than on the possibility of cooperation with the other actor ('prisoner'). Thus, the existence and the characteristics of the exchange relation are the most salient informational elements, and are prominent in the mental

model. This focusing effect is enhanced by the fact that the third party ('police') is active vis-à-vis the actor ('prisoner'), whereas the other actor is not. From a framing point of view, the relationship with the third party has two important properties that are closely related but analytically distinct. First of all, there is the *exchange label* attached to the relationship. This label likely invokes norms of competitiveness and individual gain, much like the name 'Wall Street Game' does in the PDG. In addition, since exchanges with the third party are *mutually* beneficial, *valence framing* implies that an actor perceives that striking a deal with the third party is equivalent to cooperation with the third party, rather than as defection against the other actor. Therefore, we expect actors are more likely to defect in the exchange dilemma than in the traditional 2-person 2×2 PDG of Table 1.

The second reason for expecting no cooperation in the exchange dilemma is that the third party will try to tempt each actor into defecting by making him better offers if necessary. Sally (1995: 75) in his review found that the opportunity to double one's reward by unilateral defection decreased the likelihood of cooperation by 11%-16%. Moreover, since the offers of the third party generally differ on a continuous scale, the actors' payoffs in the exchange dilemma are variable and continuous rather than fixed, as in the PDG of Table 1. This continuity of payoffs makes it easier for the third party to adapt the temptations for the actors in the exchange dilemma, without making large concessions and thereby incurring large costs. In addition to the main prediction of hardly any cooperation in the exchange dilemma, we therefore hypothesize that the third party elicits defection in the exchange dilemma by increasing the temptation payoff for the actors.

STUDY 1

The aim of Study 1 was to examine to what extent cooperation occurs between the two 'prisoners' A and C in the exchange dilemma of Table 3. The cooperation rate; i.e., the extent to which A and C do *not* exchange with B, is compared to the extent to which A and C *fail* to exchange in an exchange situation with two independent bilateral exchange opportunities, *without* externalities between A and C. In their study on bilateral exchange Dijkstra and Van Assen (2008: 28) found average rates of failure to exchange of .08 to .18 in similar exchange situations without externalities.

Study 1 contained 2 conditions: the "Dilemma" condition representing the exchange dilemma of Table 3, and the "Exchange" condition with the same exchange opportunities but without externalities. In both conditions, actors were endowed with resources and payoffs according to the third and fourth rows of

Table 2. In the Exchange condition, all the costs and benefits of exchange were private to all participants. In the Dilemma condition *negative externalities of exchange* exist between the A and the C participants, as described earlier (Table 3). To allow estimation of the effect of negative externalities per se, B was not informed of the negative externalities of the exchanges for A and C. As explained above, focusing implies that A and C in the exchange dilemma are focused on the beneficial interaction with B, and not on the negative externalities of this interaction for each other. Framing subsequently implies that the exchange with B is conceived by A and C as mutually profitable cooperation with B, instead of defection against the other actor.

If defection (i.e., exchanging with B) in the exchange dilemma is the norm and the prisoners' behavior in the Dilemma condition is not affected by the negative externalities (because of framing and focusing), we can expect a cooperation rate in the Dilemma condition similar to the rate of not exchanging in the Exchange condition, without externalities. Alternatively, if A and C in the exchange dilemma do take the negative externalities of their exchanges into account, then the proportion of not exchanging will be higher in the Dilemma than in the Exchange condition.

Because of the negative externalities of exchange in the Dilemma condition actors A and C may initially be reluctant to exchange with B. However, B can tempt A and C to exchange by making them better offers. Such tempting behavior of B is unnecessary in the Exchange condition. We test whether B tempts A and C by verifying whether A and C transfer fewer units of X and Y, respectively, in the Dilemma than in the Exchange condition.

Method

Twenty-four participants were recruited through e-mails sent to students of a Dutch university. Participants earned an average of 15 euros, and an experimental session lasted for approximately 30 minutes.

Participants were randomly assigned to roles (A, B, or C) they occupied during the entire experiment. A and C were seated opposite to B, with a divider preventing A and C from seeing each other. Study 1 had a within-subjects design, with four groups starting in the Exchange condition and the other four groups starting in the Dilemma condition. After the instructions, one practice round (without externalities) was played, after which participants played six rounds of each condition. A round ended either when no more exchanges could be made, or when 200 seconds had elapsed. After each round resources were replenished and the game started afresh. After each round subjects were told how much they had

earned that round. The value of participants' initial endowments was subtracted from their points, so that they only earned points for exchanging.

Study 1 was a face-to-face, pencil-and-paper experiment, in which participants made offers to each other through specially prepared forms. Participants used these forms to write the number of units of resource they demanded from their potential exchange partner such that all were informed of the ongoing negotiations between any pair of participants. Communication was only possible through offer forms; no speaking or gesticulation was allowed. Participants were provided with verbal instructions by the experiment leader explaining how offers could be made and accepted. Participants knew the number of rounds to be played, and the value of one resource unit for all participants in terms of points. They knew only the monetary value of a point for themselves but not for others.

Results

In the Exchange condition, the proportions of no exchange completed in six rounds across the eight groups were .083 (*SD* = .126) and .125 (*SD* = .148) for A and B, and B and C, respectively. The cooperation rates (i.e., no exchange rates) in the Dilemma condition were .188 *(SD* = .188) and .229 (*SD* = .198), respectively. Neither of the cooperation rates in the Dilemma condition was significantly larger than the corresponding no exchange rate in the Exchange condition (for A and B, $t(7) = 1.30$, one-tailed $p = .117$; for B and C, $t(7) = 1.19$, one-tailed $p = .136$). To obtain a more powerful test we combined the two exchanges by running a multilevel regression analysis on the proportions (Snijders et al. 1999). This test combining both exchanges also did not reveal a larger rate of exchange in the Dilemma (.208 [*SD* = .188]) than in the Exchange condition 1 (.104 [*SD* = .134]) ($t(16) = 1.45$, one-tailed $p = .08$). Hence we cannot reject the hypothesis that the cooperation rate (proportion of no exchange) is equal in the Dilemma and Exchange conditions. Cohen's *d* was .64, representing a medium effect size.

The observed difference in exchange proportions between the two conditions was mainly produced by one out of eight groups of participants in which A and C mutually cooperated, i.e., did not exchange with B, in the last three rounds of the Dilemma condition. Averaged across the remaining seven groups of participants in the Dilemma condition mutual cooperation occurred only in 4% of the trials, and average proportions of cooperation were .143 and .167 for A and B, and B and C, respectively.

An analysis of the average number of units of X and Y that A and C transferred to B, respectively, suggests that B did not need to convince A and C to exchange by making them better offers in the Dilemma condition. Average numbers of

units transferred by A and C respectively, were 14.5 and 34.0 in the Exchange condition, and 13.6 and 34.3 in the Dilemma condition (one-tailed Wilcoxon signed rank test; $p = .32$ and $p = .64$, respectively).

Discussion Study 1

The rate of cooperation in the Dilemma condition (.208) was close to the proportion of no exchange in the Exchange condition (.104). Moreover, this rate of cooperation is also considerably lower than the average cooperation rate found by Sally (1995) in his review (.47), and cooperation in the public goods game (about half of one's endowment; Camerer, 2003) and trust game (also about half one's endowment; Johnson and Mislin, 2011). Our current results therefore suggest that defection is the predominant action in the exchange dilemma; when a third party is present (the 'police officer') who can affect the prisoners' payoffs, the prisoners readily confess. Moreover, the mere presence of the third party sufficed to produce this effect; B did not have to make better offers to A and C in order to tempt them into a confession. Thus, the focusing of A and C on their profitable exchange opportunities with B and the framing of the situation in terms of a profitable exchange were sufficient to make A and C defect.

STUDY 2

The aim of Study 2 was to examine cooperation in the exchange dilemma as in Study 1, but with a different experimental setting that was web-based and in which two third parties were present. The cooperation rate of Study 2 is again compared to the lower benchmark observed in the Exchange condition of Study 1, i.e., .104.

The last two rows of Table 2 present the resources and payoffs the participants (A, B, C, and D) were endowed with in the exchange dilemma of Study 2. These resources and payoffs imply that profitable exchanges are possible between the connected pairs in the box network of Figure 2b. We constrained participants to a maximum of one exchange per experimental round. Just like in the Dilemma condition of Study 1, negative externalities of exchange exist between the A and the C participants: whenever A (C) exchanges with either B or D, C (A) loses 24 points. These exchange opportunities and externalities together determine the payoff matrix of Table 4 for A and C. The payoff matrix of Table 4 represents an exchange dilemma for A and C, in which there are now

two third parties (B and D).[2] Thus, this situation resembles one in which two fishermen (A and C) fishing from common waters have exchange relations with two potential buyers (B and D). Alternatively, A and C could be two country-members of the OPEC, whose exchanges in excess of the agreed quota spoil the market and harm other OPEC members.

Table 4. *The Exchange Dilemma induced by Study 2*

			C	
			Not Exchange	Exchange
A		Not Exchange	0, 0	– 24, yC – 24
		Exchange	yA – 24, – 24	yA – 48, yC – 48

Note: y_A and y_C denote the number of units of Y received by A and C respectively; $24 \leq y_A, y_C \leq 48$, for mutually profitable exchanges.

Focusing and framing in this exchange dilemma again suggest that the cooperation rate in Study 2 will equal the proportion of no exchange in an exchange situation without externalities (Exchange condition of Study 1). Again, because of the negative externalities in the exchange dilemma actors A and C may initially be reluctant to exchange with B and D. However, B and D can tempt them to exchange by letting them have more than half of the surplus of the exchange. The surplus of exchange in each relation (i.e., total payoffs without externalities) in Study 2 equals 24. To test whether B and D tempt A and C we verify whether A and C obtain on average more than 12 (disregarding externalities) from the exchanges they carry out.

Method

Sixteen participants were recruited from a participant pool at a large state university in the USA. Participants earned an average of 10 dollars. An experimental session lasted for approximately 30 minutes. Participants made offers through computer terminals, using the ExNet 3.0 software (Girard 2003).

2 In fact, Table 4 is not a dilemma situation if either A or C obtains all (48) units of Y. This is very unlikely, since it implies that B or D do not gain in their exchange with A or C. Indeed, it never occurred that A or C obtained all units of Y in Study 2.

Participants were randomly assigned to player roles (A, B, C, and D) they occupied during the entire experiment. Upon arrival they were guided to isolated experimentation booths, in which they could neither hear nor see any other participant. They then received written instructions that they kept during the entire experiment. After they finished reading the instructions three practice rounds were played, in which they played the exchange game without negative externalities between A and C. After the practice rounds it was explained to A and C how their payoffs depended on externalities and participants played ten rounds of the exchange dilemma. A round ended either when no more exchanges could be made, or when 180 seconds had elapsed. After each round resources were replenished and the game started afresh.

After each round participants were informed of their earnings in that round, but not of the earnings of other participants. The participants knew the number of rounds to be played, and the value of one resource unit for all participants in terms of points, but not the monetary value of a point for other participants. Similar to Study 1, B and D were unaware of the negative externalities between participants A and C. All participants were informed of the ongoing negotiations between any pair of participants in their group via their computer screens.

Results

The average proportion of exchanges completed in ten rounds in Study 2 was .875 for C and .9 for A. The combined cooperation rate (.112 [*SD* = .173]) was not significantly larger than the benchmark of .104 (*t*-test for two independent samples, $t(23) = .10$, one-tailed $p = ,46$)[3], with a very small effect size (Cohen's $d = .05$). Thus, we found no evidence for cooperation at all in the exchange dilemma of Study 2.

The average payoffs of A and C were not significantly larger than 12, suggesting that B and D did not need to tempt A and C into exchange. The estimated average payoffs of A and C as obtained with multilevel analysis were in fact smaller than 12 (for C 10.97, $t(36) = –1.69$, one-tailed $p = .954$; for A 10.94, $t(36) = -1.95$, one-tailed $p = .974$).

Analysis at the level of the four groups reveals the cooperation rates .3, .1, .05, and 0. No evidence for mutual cooperation was found in the latter three

3 The denominator of the t-value equals $\sqrt{.002577 + .00374}$, the first and second number representing the standard error of the proportion of Condition 1 in Experiment 1 obtained by multilevel analysis and the standard error of the combined proportion in Experiment 2, respectively. The proportions have 16 and 8 degrees of freedom.

groups. In the first group *mutual* cooperation (i.e., both A and C not exchanging in the same round) occurred only once (1%). The proportion of cooperation in the first group was .3 mainly because C exchanged in only half of the rounds. C in the first group was greedy; on average he wanted to obtain almost 90% of the profits of exchanging, a demand typically rejected by a B or D participant. His greediness, however, can also be interpreted as an attempt to instigate cooperation with A. Indeed, both A and C in the first group initially demanded very favorable exchange rates (24, which is the maximum profit) for themselves in rounds 8 to 10. This led to successful mutual cooperation in round 8, but after C was successfully tempted to exchange with a profit of 18 in rounds 9 and 10, A only obtained 16 and 8 in rounds 9 and 10, respectively. Some evidence for attempts at cooperation was found in later rounds of group 1.

Discussion Study 2

The cooperation rate in Study 2 was similar to the rate of no exchange in the Exchange condition of Study 1. Furthermore, the realized payoffs of A and C participants suggest they did not need to be tempted to exchange. Hence, similar to Study 1, the results suggest that defection is the default action in an exchange dilemma.

Although the overall results of Study 2 suggest no cooperation at all in the exchange dilemma, closer inspection of the results in one of the groups hints at attempts to cooperate. Both A and C initially made very high demands to B and D in later rounds in that group. Making such high demands renders successful exchange unlikely. In addition, the high demands provide some sort of reassurance for the participant making them. If accepted, they yield a high payoff, such that the negative consequences of the exchange of the other actor in the exchange dilemma are somewhat compensated. Nevertheless, except for one round, C and A could not resist the temptation to exchange, and the one who was exchanging second received the lowest payoff. The observation that the first exchange was more profitable than the second could have triggered A and C to try and be the first to exchange. In the end, even in the group in which attempts at cooperation were observed, the cooperation rate of .3 was low.

GENERAL DISCUSSION AND CONCLUSION

Many real-life social dilemmas involve a third party (i.e., a police officer in the language of the PDG) who is capable of changing the payoffs of the actors

in the dilemma (the 'prisoners') and is motivated to do so in order to tempt the actors into defection. We argued in this paper that these social dilemmas can be fruitfully modeled as exchange dilemmas. Moreover, based on a focusing and framing argument we predicted hardly any cooperation in the exchange dilemma.

We tested our hypotheses in two experimental studies where one (Study 1) or two (Study 2) third parties could negotiate over mutually profitable exchanges with two actors in a social dilemma. For these actors, exchanging with a third party implied defection towards the other actor. Hardly any cooperation was observed in the exchange dilemmas of both our studies; the cooperation rates of .208 (Study 1) and .112 (Study 2) were much closer to the proportions of no exchange when no externalities were present (.104) than to the average cooperation rate in PDGs found by Sally (1995) (.47).

In addition to our focusing and framing argument we also argued that in an exchange dilemma a third party can surmount any initial reluctance of actors in the dilemma to exchange by making them better offers. We did not observe this effect, however, and conclude that the focusing of the actors on their exchange opportunity with the third party and the framing of these relationships as profitable exchanges sufficed to produce very high rates of defection. There apparently was no need for third parties to tempt the actors in the social dilemma to exchange

We note that in our studies third parties were unaware of the exchange dilemma between the actors. The implication is that the 'police' (B and D) did not know that the 'prisoners' (A and C) might be reluctant, and why. Like in Tucker's anecdote, a real police officer would of course know about the dilemma between the prisoners and would therefore more readily contemplate a strategy of temptation by changing the payoffs. Since we wanted to study the effects of the externalities-induced exchange dilemma only, and not the effects of changes in information for third parties (especially when comparing exchange rates to those of the Exchange condition 1 in Study 1), we chose to keep third parties ignorant of the exchange dilemma. Our results show that temptation by the police was not necessary. Revealing the negative externalities of exchange to the third parties may have resulted in more temptation and hence higher payoffs for the actors in their exchanges with the third party, and even less cooperation than currently found in both studies.

A difference between the PDG and exchange dilemmas studied here is that play in the PDG is simultaneous, whereas it was sequential in the exchange dilemma. Our design shares features with the Alternating Prisoner's Dilemma Game (APDG) (Frean, 1994; Sally, 1995), since participants observed each other's moves. The expectation is that sequential play renders cooperation *more* likely (Hayashi et al. 1999; Kiyonari et al. 2000). The intuition is that C will defect after

observing a defection of A in the APDG, resulting in a game where the T payoff is unattainable for A, thereby making defection less and cooperation more likely in the APDG. The fact that we nonetheless found hardly any cooperation in the exchange dilemma emphasizes the robustness of our findings.

Studies 1 and 2 were different in a number of important aspects. Study 1 was a face-to-face paper-and-pencil study, whereas Study 2 was computer-mediated with participants unable to communicate but through sending and receiving offers via their computer terminals. In Study 1 six rounds of exchange were played with only one third party, while in Study 2 ten rounds were played with two third parties. Finally, in Study 1 the actors in the dilemma (i.e., the A and C subjects) were unequal with respect to their endowments and payoffs, whereas in Study 2 the actors were identical in these respects. Despite all these differences, results of the exchange dilemmas in both studies were strikingly similar: hardly any cooperation and no need for temptation by third parties. This makes the evidence that the high rates of defection in the exchange dilemma were indeed caused by the addition of the social exchange relations with the third parties more persuasive. Moreover, it also fortifies our argument that in still other contexts, such as in the real-life exchange dilemmas described in the introduction, the mechanism will operate similarly.

Our results suggest that the fundamental problem of cooperation in many real-life social dilemmas may be more severe and harder to solve than suggested by traditional Prisoner's Dilemma research due to the presence of third parties. That is, cooperation will be harder to attain and maintain in the presence of an active third party interested in defection, than without such a party. In such cases, an understanding of the relationship between the agents in the dilemma with the third party must be an integral part of the explanation of observed instances of cooperation and defection.

Important elements of the relationships between the agents and the third party include the possibilities of reputation formation (e.g., Raub et al. 1990) and punishment (e.g., Fehr et al. 2002) among the prisoners. Both these factors are known to have a positive effect on cooperation in traditional PDGs (Rand et al. 2013). It is an open question to what extent the positive effects of reputation formation and punishment can override the temptation of the actors in the dilemma to exchange with the third party, thereby increasing the cooperation in exchange dilemmas.

The crucial element is obviously the third party. We suspect that many features of the relation with the third party have a strong effect on cooperation in the exchange dilemma. First, the presence of multiple third parties competing for exchange with the actors in the dilemma will increase the temptation to exchange for these actors, further decreasing cooperation. This effect is aggravated by

the fact that the negative externalities of the exchange become smaller when temptation increases. Another factor is the access of the third parties to the actors. We expect that if the third parties can directly negotiate with the actors in the dilemma, then cooperation will be more difficult. Examining the role of the third party and its relation with the actors in an exchange dilemma hence provides interesting and important avenues for future research.

REFERENCES

Axelrod, Robert (1984), *The Evolution of* Cooperation, New York, NY: Basic.

Buchan, Nancy R. – Rachel T.A., Croson – Robin M. Dawes (2002), "Swift neighbors and persistent strangers: A cross-cultural investigation of trust and reciprocity in social exchange", *American Journal of Sociology* Vol. 108, No 1, pp. 168-206. DOI: 10.1086/344546

Camerer, Collin (2003), *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Cook, Karen S. – Richard M. Emerson – Mary R. Gillmore – Toshio Yamagishi (1983), "The distribution of power in exchange networks: Theory and experimental results", *American Journal of Sociology* Vol. 89, No 2, pp. 275-305.

Dawes, Robin M (1980), "Social dilemmas", *Annual Review of Sociology* Vol. 31, pp. 187-276. DOI: 10.1146/annurev.ps.31.020180.001125

Dijkstra, Jacob (2009), "Externalities in exchange networks: An adaptation of existing theories of exchange networks", *Rationality and Society* Vol. 21, No 4, pp. 395-427. DOI: 10.1177/1043463109335612

Dijkstra, Jacob – Marcel A.L.M. van Assen (2008), "Transferring goods or splitting a resource: Testing the consequences of the violation of a basic assumption in exchange research", *Social Psychology Quarterly* Vol. 71, No 1, pp. 17-36.

Dijkstra, Jacob – Marcel A.L.M. van Assen – Frans N. Stokman (2008), "Outcomes of collective decisions with externalities predicted.", *Journal of Theoretical Politics* Vol. 20, No 4, pp. 415-441. DOI: 10.1177/095162980809377

Dufwenberg, Martin – Simon Gächter – Heike Hennig-Schmidt (2011), "The framing of games and the psychology of play", *Games and Economic Behavior* Vol. 73, No 2, pp. 459-478. doi:10.1016/j.geb.2011.02.003

Fehr, Ernst – Simon Gächter (2000), "Fairness and Retaliation: The Economics of Reciprocity", *Journal of Economic Perspectives* 14 (3): 159–82.

Fehr, Ernst – Urs Fischbacher (2003), "The nature of human altruism", *Nature* Vol. 425, No 6960, pp. 785-791. doi:10.1038/nature02043

Fehr, Ernst – Simon Gächter (2002), "Altruistic punishment in humans", *Nature* Vol. 415, pp. 137-140. doi:10.1038/415137a

Frean, Marcus R (1994), "The prisoner's dilemma without synchrony", *Proceedings: Biological Sciences* Vol. 257, No 1348, pp. 75-79. DOI: 10.1098/rspb.1994.0096

Girard, C.D (2003), "A digital library for social science research software", Retrieved February 25, 2011 (http://www.allacademic.com/meta/p107695_index.html).

Hardin, Garrett (1968), "The tragedy of the commons" *Science* Vol. 162, pp. 1243-1248.

Hayashi, Nahoko – Elinor Ostrom – James Walker – Toshio Yamagishi (1999), "Reciprocity, trust, and the sense of control: A cross-societal study", *Rationality and Society* Vol. 11, No 1, pp. 27-46.

Johnson, Noel, D – Alexandra A. Mislin (2011), 'Trust Games: a Meta-Analysis', *Journal of Economic Psychology*, 32: 865-89.

Johnson-Laird, Phillip N (1983), *Mental models: Towards a cognitive science of language, inference and consciousness*, Cambridge: Cambridge University Press.

Kahneman, Daniel – Amos Tversky (1979), "Prospect theory: An analysis of decision under risk", *Econometrica* Vol. 47, No 2, pp. 263-291.

Kanazawa, Yusuke (2007), "The promotion and evolution of cooperation through projection: Implications for social dilemmas and trust", *Journal of Mathematical Sociology* Vol. 31, No 2, pp. 187-204. Doi: 10.1080/00222500601188502

Kollock, Peter (1998), "Social dilemmas: The anatomy of cooperation", *Annual Review of Sociology* Vol. 24, pp. 183-214.

Kiyonari, Toko – Shogehito Tanida – Toshio Yamagishi (2000), "Social exchange and reciprocity: Confusion or a heuristic?", *Evolution and Human Behavior* Vol. 21, No 6, pp. 411-427. doi:10.1016/S1090-5138(00)00055-6

Legrenzi, P. – V. Girotto – P. N. Johnson-Laird (1993), "Focussing in reasoning and decision making", *Cognition* Vol. 49, No 1-2, pp. 37-66. doi:10.1016/0010-0277(93)90035-T

Levin, Irwin P. – Sandra L. Schneider – Gary J. Gaeth (1998), "All frames are not created equal: A typology and critical analysis of framing effects", *Organizational Behavior and Human Decision Processes* Vol. 76, No 2, pp. 149-188. doi:10.1006/obhd.1998.2804

Liberman, Varda – Steven M. Samuels – Lee Ross (2004), "The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves", *Personality and Social Psychology Bulletin* Vol. 30, No 9, pp. 1175-1185. DOI: 10.1177/0146167204264004

Macy, Michael W (1991), "Learning to cooperate: stochastic and tacit collusion in social exchange", *American Journal of Sociology* Vol. 97, No 3, pp. 808-843.

Molm, Linda, D. (1997), *Coercive power in social exchange*, Cambridge: Cambridge University Press.

Nowak, Martin – Karl Sigmund (1993), "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game", *Nature* Vol. 364, pp. 56-58. doi:10.1038/364056a0

Nowak, Martin – Karl Sigmund (2005), "Evolution of indirect reciprocity", *Nature* Vol.437, pp. 1291-1298. doi:10.1038/nature04131

Ostrom, Elinor (1990), *Governing the commons: The evolution of institutions for collective action*, Cambridge University Press.

Poundstone, William (1992), *Prisoner's dilemma*, New York, NY: Bantam Doubleday Dell Publishing Group.

Rand, David G – Martin A. Nowak (2013), "Human cooperation", *Trends in Cognitive Sciences* Vol. 17, No 8, pp. 413-425. doi:10.1016/j.tics.2013.06.003

Raub, Werner – Jeroen Weesie (1990), "Reputation and efficiency in social interactions: An example of network effects", *American Journal of Sociology* Vol. 96, No 3, pp. 626-654.

Sally, David (1995), "Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992" *Rationality and Society* Vol. 7, No 1 pp. 58-92. doi: 10.1177/1043463195007001004

Snijders, Tom A.B. – Roel J. Bosker (1999), *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, London: Sage.

Snijders, Chris – Gideon Keren (2001), "Do You Trust? Whom Do You Trust? When Do You Trust?", *Advances in Group Processes* 18: 129–60.

Takahashi, Noboyuki (2000), "The emergence of generalized exchange", *American Journal of Sociology* Vol. 105, No 4, pp. 1105-1134.

Thaler, Richard H (1988), "Anomalies: The Ultimatum Game", *The Journal of Economic Perspectives* 2 (4): 195–206.

Thomson, Robert -Frans N. Stokman – Christopher H. Achen – Thomas König eds. (2006), *The European Union decides*, Cambridge: Cambridge University Press.

Tversky, Amos – Daniel Kahneman (1981), "The framing of decisions and the psychology of choice", *Science* Vol. 211, pp. 453-458. DOI: *10.1126/science.7455683*

Udehn, Lars (1996), *The limits of public choice: A sociological critique of the economic theory of politics*, London: Routledge.

Van Assen, Marcel A.L.M – Frans N. Stokman – Reinier C.H. Van Oosten (2003), "Conflict measures in cooperative exchange models of collective

decision making", *Rationality and Society* Vol. 15, pp. 87-115. doi: 10.1177/1043463103015001072
Willer, David (1999), *Network exchange theory*, Westport, CT: Praeger Press.
Willer, Robb (2009), "Groups reward individual sacrifice: The status solution to the collective action problem", *American Sociological Review* Vol. 74, No 1, pp. 23-43. doi: 10.1177/000312240907400102