Journal of Economic Behavior & Organization
Vol. 47 (2002) 275–290

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Identifying cooperative behavior: some experimental results in a prisoner's dilemma game

Jeannette Brosig

Faculty of Economics and Management, Otto-von-Guericke-Universität Magdeburg, P.O. Box 4120, D-39106 Magdeburg, Germany

Received 30 March 1999; received in revised form 18 May 2001; accepted 29 May 2001

Abstract

In recent years, experimental economists have discovered that people exhibit different patterns of cooperative behavior. This paper presents findings from a face-to-face experiment that analyzed whether individuals who possess a willingness to cooperate can credibly signal it and whether it is recognizable by the partner. Results revealed that both capabilities, signaling and recognizing, depend upon the individual's propensity to cooperate. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C91; C78

Keywords: Experimental economics; Communication; Cooperation; Fairness

1. Introduction

The investigation into the cooperative behavior exhibited by individuals constitutes a major research topic in experimental economics. Cooperative behavior is manifested whenever individuals maximize the joint payoff, in lieu of their own payoff, in prisoner's dilemma and public good games. When observed in these games, this behavior attracts attention for it runs contrary to the more self-centered behavior predicted by standard game theory. For this reason, an extensive literature has developed attempting to explain this behavioral pattern. Various divergent theoretical explanations have been proposed, including "pure" as well as "warm glow" altruism (Andreoni, 1993; Andreoni and Miller, 1995; Offerman et al., 1996), the relative payoff position of individuals (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), "cooperative gain seeking" (Brandts and Schram, 1996), different sorts of reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 1998; Falk and Fischbacher, 1998), "quasi-maximin preferences" (Charness and Rabin, 2000), and the combination of altruism with error effects (Palfrey and Prisbrey, 1997; Anderson et al., 1998).

E-mail address: jeannette.brosig@ww.uni-magdeburg.de (J. Brosig).

PII: S0167-2681(01)00211-6

The majority of theories that seek to explain cooperative behavior postulate the existence of different types of individuals, each type exhibiting a distinctive pattern of behavior. The existence of these different behavioral types has been confirmed by experimental findings. For example, in a finitely repeated one-shot prisoner's dilemma game, Andreoni and Miller (1993) observed that some subjects always chose to cooperate with their partners while others either chose not to cooperate or played a mixed strategy. Similar results were reported by Cooper et al. (1996). In public good experiments, Weimann (1994) systematically identified three types of subjects. Each of these types made differing decisions regarding the quantity and timing of the contributions made. The same types of individuals were observed in the public good experiments conducted by Isaac et al. (1994).[1] Analyzing behavior in a centipede game, McKelvey and Palfrey (1992) also concluded that there existed a small fraction of subjects who decided to cooperate throughout the game while others did not.

Recently, evolutionary game theory has been developed into a helpful tool for explaining the existence of these different types of individuals. According to this theory, only those behavioral patterns that produce a relative fitness advantage for an individual will survive the selection process (i.e. those behavioral patterns that are reproductively successful).[2] Based on this theory, it seems very unlikely that cooperation would be a selected behavioral pattern, because it enhances the fitness of non-cooperative individuals rather than assisting the cooperative ones. This is only true, however, as long as no information is available concerning the types of individuals being confronted with that would allow for selected interactions to occur.[3] In order to justify the existence of cooperative types, evolutionary models assume that relevant information can be communicated in different ways. In particular, some of these models postulate the existence of signals indicating a partner's type, and thus, offer explanations for cooperative behavior even in one-shot situations (Robson, 1990; Yang, 1993; Amann and Yang, 1998).[4] These signals are only useful, however, if they can be communicated successfully. This makes communication a necessary condition for the evolution of one-shot cooperation.[5]

[1] This conclusion is based on my own calculations utilizing data graciously provided by James M. Walker.

[2] Evolutionary game theory uses the equilibrium concept of evolutionary stability that defines only those strategies as evolutionary stable that have a higher reproductive success rate than any other mutant strategy that might arise. For a more detailed description, see Weibull (1995). Hammerstein and Selten (1994) provide an excellent overview of the biological applications. An introduction into the rapidly developing field of evolutionary psychology is given by Barkow et al. (1992), and Buss (1999).

[3] In a bargaining experiment, Charness (2000) showed that providing information about the allocations chosen by subjects in a dictator game increased the payoff to those, who were more generous dictators. Cain (1998) obtained similar results in a prisoner's dilemma game after subjects were informed about their partner's chosen dictator allocation. Interestingly, in our research, fair behavior in the dictator game is per se not related to joint payoff maximization.

[4] Other models that explain the evolution of cooperation are based upon the assumption of repeated interaction using some form of reciprocity (see for e.g. Kreps et al., 1982; Guttman, 1996, as well as the learning model proposed by Vogt, 2000).

[5] The fact that communication has a considerable influence on cooperative behavior has been reported in various experimental studies. In prisoner's dilemma games, Dawes et al. (1977), Orbell et al. (1988), as well as Bohnet and Frey (1995) found that pre-play communication enhanced the number of decisions to cooperate. Similar observations were reported in experiments on common pool resource games (Ostrom and Walker, 1991; Ostrom et al., 1994; Hackett et al., 1994). In public good games, Isaac et al. (1985), Isaac and Walker (1988), and Brosig

There is, however, the fundamental question of whether cooperative individuals, who are offered the opportunity to communicate, are able to signal their willingness to cooperate (i.e. their willingness to choose actions that maximize the joint payoff). According to Frank (1988), such individuals are endowed with an advanced emotional system that not only provides the motivation for the cooperative behavior, which they exhibit even in anonymous situations, but also enables them to signal their propensity to cooperate. Specifically, he assumes that cooperative individuals experience strong emotions, such as sympathy, compassion, guilt, and shame. These emotions prevent them from defecting and they can be communicated via signals that are partially independent of their direct control, thus making them difficult to imitate. In a formal model, Frank (1987) considered a population whose members, consisting of both cooperators and defectors, could engage in a joint venture that allowed them to maximize a joint payoff. The joint venture was modeled as a one-shot prisoner's dilemma game. While cooperators disliked cheating and were able to signal this attitude with a certain degree of credibility, defectors never chose cooperation (i.e. they maximized their own payoff instead of the joint payoff). Frank (1987) was able to demonstrate the existence of an evolutionary stable outcome with cooperation. Harrington (1989) observed, however, that this result depends strongly on how reliably the willingness to cooperate can be signaled and identified by the partner. In a later experiment, Frank et al. (1993) tested this assumption by asking subjects to predict the decisions of their partners in one-shot prisoner's dilemma games. According to their findings, the propensity to cooperate seems to be discernible by others. There are, however, some aspects of their experimental design, described in greater detail later in the paper, that shed some doubt on this conclusion. For example, explicit statements regarding the willingness to defect in one-shot prisoner's dilemma games are per se very credible. Therefore, the inclusion of these statements in the analysis might have biased the findings of Frank et al. (1993).[6] In this paper, an experimental design is presented that not only avoids the shortcomings in the design used by Frank et al. (1993), but also allows for the testing of some further hypotheses.

The existence of cooperative individuals, according to Frank (1988), can be explained by certain emotional dispositions that help them solve the commitment problem. Emotions are essential, therefore, for signaling an individual's behavioral type. It is easy to imagine then, that emotions also play an important role in recognizing these signals, because emotional traits, such as sensibility and empathy, are required for discerning a cooperative disposition in others. Moreover, knowing one's own emotions and being sensitive to them should make it easier to identify those signals that characterize these same emotional traits in others. The ability to recognize that others have a willingness to cooperate, therefore, should be attributed to an individual's emotional make up in the same manner as is the cooperative disposition itself.[7] As a result, cooperative individuals might be generally better at predicting the cooperativeness of others than defecting ones. Frank (1987) gave a

et al. (2000) also observed the cooperation enhancing effect of communication. For an overview of communication experiments, see Sally (1995) or Ledyard (1995).

[6] For example, in a bargaining experiment, Ockenfels and Selten (2000) found no evidence that there existed external clues to the emotional state.

[7] This suggestion conforms with recent findings by neuroscientists, who have found that the capability to recognize the emotion fear from facial expressions is associated with neocortical regions, located in the right hemisphere of the brain, that is also involved in emotional expression and experience (see for e.g. Adolphs et al., 1996).

reasonable explanation for why there might have evolved such a close relationship between the capability to recognize cooperation and the cooperative disposition itself. He assumed that there exist some costs associated with developing the ability to scrutinize others (i.e. the effort that must be spent in acquiring the required sensibility for identifying the other's disposition). As Frank (1987) demonstrated, it would only be in the interest of cooperative individuals to bear these costs, and thus, to become sensitized. Taking these considerations into account, the following working hypotheses formed the framework for the experiment.

Hypothesis 1. Individuals differ according to the extent of their cooperative behavior in anonymous interactions.

With regard to cooperative behavior, Frank (1988) postulated that there exist two types of individuals: cooperative individuals, who are assumed to cooperate in anonymity (i.e. those individuals who always decide to maximize the joint payoff in an anonymous interaction), [8] and defecting individuals, who are assumed always to defect (i.e. those individuals who always decide to maximize their own payoff). We would expect, therefore, to observe both types of behavioral patterns when testing individuals for cooperative behavior in anonymity.

Furthermore, cooperative individuals, those individuals who decide for joint payoff maximization in anonymity, are expected to be better at recognizing the other individual's behavioral type. This means that they should be better able to identify the other individual's willingness to cooperate in the experiment. Thus, our second hypothesis follows.

Hypothesis 2. Individuals who choose to cooperate in anonymity are better at identifying their partner's willingness to cooperate in a prisoner's dilemma game than individuals who choose to defect.

If it is assumed that individuals with a cooperative disposition experience strong emotions that prevent them from cheating, then it should also be observed that, after communicating, these individuals will not exploit the expected cooperation of their partner in prisoner's dilemma games. Consequently, our third hypothesis occurs.

Hypothesis 3. Individuals who choose to cooperate in anonymity are less inclined to exploit the predicted cooperation of other individuals in a one-shot prisoner's dilemma game.

This hypothesis leads to the analysis of the related question: will those individuals, who choose to cooperate in anonymity, also decide to cooperate even after they have recognized that their partner will defect? In Frank's (1987) model, after recognizing that their partner is a defector, cooperative individuals are assumed to stop cooperating and to take an outside option where they receive the same payoff as in the case of mutual defection. This gives rise to the conclusion that cooperative individuals are not "unconditional cooperators", who find pleasure in cooperating per se, but rather "conditional cooperators", who choose to cooperate

[8] In particular, we refer to anonymous joint ventures where cooperative individuals are forced to participate and which are, therefore, similar to settings where cooperators and defectors look alike (see Frank, 1988, p. 58).

as long as their partner is not clearly identified as defecting.[9] Frank (1987), however, does not make any explicit assumptions about how cooperators will decide after recognizing that their partner is a defector and they are forced to participate in the joint venture (i.e. when they do not have an outside option). We have made no assumptions in this regard either, but simply experimentally test in order to determine whether cooperation is pursued even when cooperative individuals receive a clear signal concerning their partner's intention to defect.

Cooperation is defined in this paper as a form of other-regarding behavior. A form that is aimed at joint payoff maximization and, as a result, at the realization of an efficiency gain. There are, however, other forms of other-regarding behavior that exclusively refer to pure distributional concerns, such as the fair behavior observed in the dictator game (Kahneman et al., 1986; Forsythe et al., 1994). Is there a relationship between these two kinds of other-regarding behavior? This question relates directly to Frank's (1988) proposition that his model could also be applied to the evolution of fair behavior, suggesting that both an individual's cooperative behavior and fair behavior can be attributed to the same emotional disposition and are, therefore, somehow related to each other.[10] Following this proposition, we would expect both kinds of other-regarding behavior to reveal similar characteristics. Specifically, our three hypotheses regarding cooperative behavior should also hold true with regard to fair behavior. Therefore, the three hypotheses stated below were considered in addition.

Hypothesis 1a. Individuals differ according to the extent of their fair behavior in anonymous interactions.

Hypothesis 2a. Individuals who choose a fairer distribution in anonymity are better at identifying their partner's willingness to cooperate in a prisoner's dilemma game than individuals who choose a less fair one.

Hypothesis 3a. Individuals who choose a fairer distribution in anonymity are less inclined to exploit the predicted cooperation of other individuals in a one-shot prisoner's dilemma game.

2. Experimental design

In order to test whether individuals have the ability to signal a willingness to cooperate and to recognize these signals from others, a two part experimental design was employed. In the first part, the classification part, two procedures were used to identify subjects according to their other-regarding behavior. The decomposed game technique was employed to test subjects for cooperative behavior and the dictator game was employed to test subjects for fair behavior. In the second part, the main part, they played a prisoner's dilemma game after being allowed to communicate with each other.

---

[9] Different versions of "conditional cooperation" have been modeled by Rabin (1993), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), as well as by Brandts and Schram (1996).

[10] See Frank (1988), p. 168.

A total of 24 sessions were conducted (including one pilot session), utilizing 143 students at the Otto-von-Guericke-Universität Magdeburg. [11] At the beginning of each session, subjects were led individually into separate rooms where they drew an identification number and received written instructions. [12] This procedure ensured that, during the classification part of the experiment, subjects could not form any attitudes regarding their potential partners (i.e. according to the hypotheses, identify their disposition).

The decomposed game technique, employed as one of the two classification procedures, was developed by social psychologists and attempts to assess an individual's "social motivation". [13] Using this technique, subjects played 24 decomposed games. In each of these games, they were asked to choose between two "own-other" payoff combinations. Each of these payoff combinations assigned a certain amount of money to the subjects themselves, the own payoff x, and a certain amount to the other subject, the other payoff y. The monetary values of these payoffs were determined such that when plotted as ordered pairs (x, y), in a two-dimensional own-other payoff space, they would be located at 24 equally spaced points around a circle, centered at the origin (0, 0) and with an arbitrary radius of 15 monetary units. Subjects were asked to choose between two own-other payoff combinations that were located at two adjacent points around this circle. This implies that $x^2 + y^2 = 15^2$ and that $x + y$ is not a constant monetary sum. [14]

The pairs of subjects playing these games remained unchanged throughout the whole classification procedure. Thus, the payoffs received by subjects were determined by the 24 decisions that they made themselves and by the 24 decisions that were made by their partners. Subjects were informed of this, but in order to avoid strategic considerations, they did not get any information concerning the identities of their partners nor did they receive any feedback concerning the decisions made by them. Adding up the amounts of money chosen by subjects for themselves and for their partners separately, an estimate of the importance given by the subject to the own payoff and to the partner's payoff was obtained. These estimates were used to approximate a vector in the own-other payoff space, representing the individual's type. Using a standard classification procedure developed for this technique, subjects with a vector lying between 67.5 and 112.5° were classified as "altruistic" (i.e. maximization of other's payoff); with a vector between 22.5 and 67.5° as "cooperative" (i.e. maximization of the sum of own and other's payoff); with a vector between 0 and 22.5° or between 337.5 and 360° as "individualistic" (i.e. maximization of own payoff); and with a vector between 292.5 and 337.5° as "competitive" (i.e. maximization of the difference between own and other's payoff). The length of this vector serves as an index for a subject's consistency (i.e.

[11] One session was conducted with five instead of six subjects, and thus, two instead of three pairs.

[12] All additional questions were answered by the same experimenter. The instructions stated that nobody would receive any information concerning their partner's decision, identity, or about the decisions made by others. Furthermore, subjects were informed that the experiment consisted of three independent games, but they were left ignorant concerning the content of each game. The instructions are available from the author upon request.

[13] See Griesinger and Livingston (1973). The version employed in our experiment was developed by Liebrand (1984) and McClintock and Liebrand (1988). A more detailed description of this technique is provided in the Appendix A.

[14] The monetary units used were called Labdollars (LD). Having finished the decomposed games, the dictator games and the prisoner's dilemma games, subjects were paid off in German marks (DM) at a rate of 1 LD = 0.25 DM.

it indicates whether the chosen own-other payoff combinations are the closest ones to the subject's motivational vector). Making 24 consistent choices yields a vector length of 30 (twice the radius of the circle). Random choices, on the other hand, result in a vector length of zero. Subjects were classified only if the length of their vector exceeded 7.5 (one-fourth of the maximum vector).

In the dictator game (see Kahneman et al., 1986; Forsythe et al., 1994)—the other classification procedure—subjects were asked to divide 30 LD between themselves and an anonymous recipient. They were explicitly informed that their anonymous recipients are participants in a similar future experiment, who would not be given the opportunity to make an allocation of this type. Subjects were classified according to the amounts of money they allocated: subjects who gave less than 10 LD to recipients were labeled "self-centered" and those who gave 10 LD or more were labeled "beneficent". [15]

In order to test whether a relationship existed between these two kinds of other-regarding behavior measured by the classification procedures, it was necessary to recognize that they, basically, differed with regard to two significant factors.

(a) Realization of mutual benefits: While subjects in the decomposed games had the opportunity to maximize the joint payoff and, as a result, to realize a mutual benefit, in the dictator game subjects were asked to divide a constant sum of money between themselves and an anonymous recipient.
(b) Interdependency of decisions: While in the decomposed games the final outcome depended upon the decisions of both subjects, in the dictator game each subject determined the final outcome alone.

In order to control for sequential interdependencies between the decomposed games and the dictator game, in 12 of the 24 sessions subjects first played the dictator game and then the decomposed games. In the other 12 sessions, the sequence was reversed.

In the main part of the experiment, subjects played a one-shot two-person prisoner's dilemma game. Before making their decision, subjects were given the opportunity to communicate with their partners in a separate room for a maximum of 10 min. No restrictions were applied to the content of these discussions. Subjects were allowed to talk freely about the prisoner's dilemma and to make non-binding promises concerning their decisions. In contrast to the experiment conducted by Frank et al. (1993), no one else took part in these discussions. As a result, external influences on the communication itself and possibly on the predictions that were made were avoided. [16] In order to analyze the content and the structure of these conversations, two-thirds of them were videotaped. After communicating, subjects were given a questionnaire, which they had to complete in separate rooms. In this questionnaire, subjects were asked to specify their own decision within the prisoner's

[15] This classification ensures that subjects labeled as beneficent are identical with those, whose allocations are the closest to the "fair" split of 15 LD for each. Furthermore, this allocation has the property that it can be approximated by a vector of about 22.5°. It is, therefore, in a sense similar to the demarcation line between cooperative and individualistic subjects in the decomposed games.

[16] There was one additional person in each room to videotape the discussion and/or to ensure that the communication lasted no longer than 10 min. This person was not allowed, however, to join in the discussion.

dilemma and to predict the decision of their partner. The content of this questionnaire was disclosed to the subjects before the communication period began.

In testing the hypotheses, it was extremely important to guarantee the anonymity of the decisions made in the prisoner's dilemma games. Specifically, the possibility that subjects could infer their partner's decision from the payments that they received needed to be avoided. [17] This is why the main part of the experiment was repeated three times, each time using different partners. One of the three prisoner's dilemma games played was chosen randomly to determine the payoffs for each of the subjects. Due to the fact that subjects were not told which game had been selected, they could not subsequently determine the decisions that had been made by their partners.

3. Results

3.1. Overall accuracy of the predictions

Frank et al. (1993) discovered that an individual's tendency to cooperate or to defect was discernible by others. For that reason, we were interested in determining whether our data would confirm this result, despite the heretofore mentioned differences in the experimental designs as well as differences in our subject selection criteria. In contrast to Frank et al. (1993), we not only excluded subjects from the sample who personally knew their partners before communicating, but we also excluded those subjects whose partners had explicitly stated during the communication period that they would decide to defect. These subjects were excluded because such an announcement is a very credible signal of future defection. [18] Based on these selection criteria, 14 subjects were excluded from the first round. Fig. 1 displays, in percentage terms, the accuracy of the first round predictions ($N = 129$) that were made by subjects playing the prisoner's dilemma games.

The predictions made by subjects regarding their partner's decision achieved an overall accuracy rate of 67 percent, which is higher than the expected accuracy rate of 59 percent reached by chance alone. [19] The hypothesis that the predictions made by subjects are independent of the decisions subsequently made by their partners is clearly rejected ($P = 0.023$; χ^2-test). Even with the more stringent restrictions that we placed on our experimental method, the findings of Frank et al. (1993) can be confirmed. Subjects are indeed able to predict their partner's decision with considerable accuracy. [20]

[17] The procedure used by Frank et al. (1993) might have distorted the incentives in the prisoner's dilemma game.

[18] With regard to the one-third of the communication periods that were not videotaped, we could infer from the questionnaires whether the intention to defect had been explicitly announced. The experimental data are available from the author upon request.

[19] The expected accuracy rate was determined on the basis of the overall rates predicted for cooperation and defection and the corresponding actual rates: $(65.1)(0.798) + (34.9)(0.202) = 59.0$ percent. This computational method was also used by Dawes et al. (1977) and Frank et al. (1993). Unless otherwise indicated, the statistical tests reported are two-tailed with a 5 percent level of significance.

[20] Interestingly, the predictions observed by Frank (1988) were significantly less accurate when the pre-play communication was restricted to an irrelevant topic.

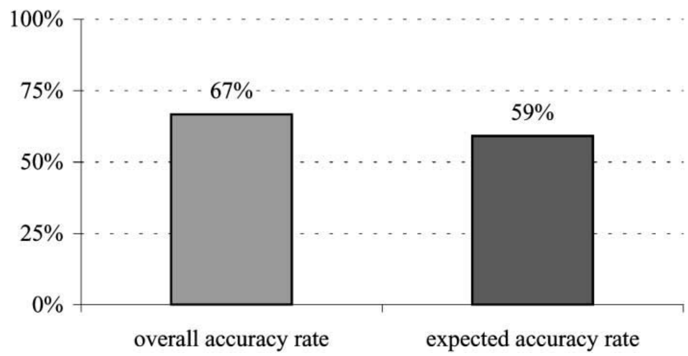

Fig. 1. First round predictions: actual vs. expected accuracy.

3.2. Classification by the decomposed game technique

In the decomposed games, 142 out of 143 subjects could be classified according to their behavioral type (it was impossible to classify one subject who appeared to be choosing randomly with a 3 percent index of consistency). On average, the choices made by the subjects yielded an overall index of consistency of 89 percent. According to the classification scheme used: 2 subjects (1.4 percent) were labeled as altruistic, 53 subjects (37.3 percent) as cooperative, 85 subjects (59.9 percent) as individualistic, and 2 subjects (1.4 percent) as competitive. Thus, Hypothesis 1 stating that individuals differ according to their cooperative behavior in anonymous interactions can be confirmed.

Analyzing the quality of the first round predictions made by cooperative and individualistic subjects, we observe that the cooperative subjects made a slightly better prediction of their partner's decision than did the individualistic subjects.[21] Thus, the difference between the overall accuracy rate and the expected accuracy rate is higher for cooperative than for individualistic subjects (see Fig. 2).

While the difference between the overall accuracy rates of cooperative and individualistic subjects is not significant ($P = 0.183$; binomial test, one-tailed), when considering only the cooperative subjects the hypothesis that predictions made by these subjects are independent from the decisions subsequently made by their partners is rejected (cooperative subjects: $P = 0.049$, individualistic subjects: $P = 0.466$; χ^2-test).[22] This implies that only the cooperative subjects predicted the behavior of their partners significantly better than chance. This observation supports Hypothesis 2.

[21] The two subjects classified as altruistic were assigned to the cooperative group and the two subjects classified as competitive were assigned to the individualistic group.

[22] The basic impression received from our data in the first round remain unaltered if the analysis is extended to include all three rounds. Although, with regard to the average accuracy rate of each subject's predictions in all three rounds, we also found no significant difference between cooperative and individualistic subjects ($P = 0.221$; Mann–Whitney U-test). For all analyses of Hypotheses 2 and 3, we used the same sample (44 cooperative and 67 individualistic subjects ($N = 111$)). That is, we excluded all subjects from the sample who personally knew one of their partners in the three rounds or whose partner had explicitly stated an intention to defect in one of the rounds.

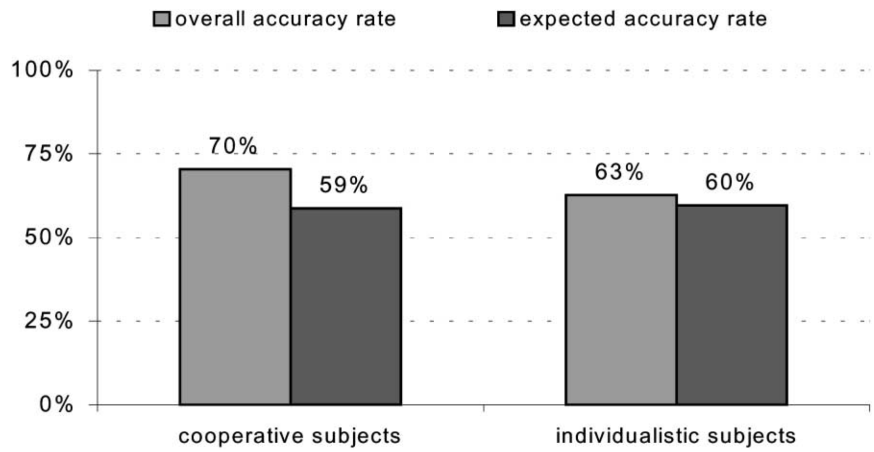

Fig. 2. First round predictions: cooperative vs. individualistic subjects.

In order to analyze Hypothesis 3, all subjects were grouped according to their predictions and their own decisions. Table 1 displays this breakdown by behavioral type and according to the number of rounds played.

Table 1 demonstrates that even under conditions of anonymity and independent of behavioral type, most subjects did not exploit their partners when they expected them to play cooperatively. Thus, when subjects believed that their partner would cooperate, they hesitated to maximize their own payoff through defection. This gives rise to the assumption that cooperation is not solely motivated by strategic considerations. Analyzing cooperative and individualistic subjects separately, we found no significant difference in the observed frequency of exploitation (first round: $P = 0.166$; binomial test, one-tailed; average exploitation in all three rounds: $P = 0.132$; Mann–Whitney U-test). According to these results, after communication, not only cooperative but also individualistic subjects were only slightly inclined to exploit their partner's anticipated cooperation. Moreover, when examining the behavior of cooperative subjects in isolation, over 90 percent of those, who predicted that their partner would defect, also chose to defect. This indicates, that after

Table 1
Behavior in prisoner's dilemma games: cooperative and individualistic subjects

Round	Classification	Prediction: cooperation		Prediction: defection	
		Decision: cooperation	Total	Decision: cooperation	Total
1	Cooperative	29 (85.3 percent)	34 (100 percent)	1 (10.0 percent)	10 (100 percent)
	Individualistic	43 (76.8 percent)	56 (100 percent)	0 (0.0 percent)	11 (100 percent)
1–3	Cooperative	94 (87.8 percent)	107 (100 percent)	2 (8.0 percent)	25 (100 percent)
	Individualistic	106 (75.2 percent)	141 (100 percent)	6 (10.0 percent)	60 (100 percent)

communication, cooperation only resulted when the partner was clearly identified as also cooperating.

Altogether, the results stemming from the classification by the decomposed game technique seem to support Hypothesis 2, but do not confirm Hypothesis 3. [23]

3.3. Classification by the dictator game

Analyzing fair behavior in the dictator game, we found that the order in which the two classification procedures were administered exerted a significant influence on the allocation decisions that were subsequently made by subjects ($P = 0.050$; Mann–Whitney *U*-test). [24] The 71 subjects who played the dictator game first (DIG 1) gave on average 8.0 LD to the anonymous recipient. The 66 subjects who filled in the questionnaire for the decomposed games first (DIG 2), however, subsequently allocated the lesser amount of 6.1 LD on average, 1.9 LD less. Noting this difference, we examined both treatments separately. According to our classification scheme 61 percent of the DIG 1 subjects (43) were classified as beneficent and 39 percent as self-centered (28). In contrast only 45 percent of the DIG 2 subjects (30) were classified as beneficent, 16 percent less, and 55 percent as self-centered (36). [25] Nevertheless, Hypothesis 1a is supported by both treatments. Individuals differ according to the extent of their fair behavior in anonymous interactions.

This result made us wonder if a relationship existed between the decisions made in the dictator game and those made in the decomposed games. The Spearman rank-order correlation coefficients, however, relating the vectors in the decomposed games to the amounts of money allocated in the dictator game were not significant (DIG 1: $\rho = 0.128$, $P = 0.2802$; DIG 2: $\rho = 0.126$, $P = 0.3124$). Obviously, the allocations chosen in the dictator game revealed quite different behavioral characteristics than the decisions made regarding the own-other payoff combinations in the decomposed games. This suggested that the two kinds of other-regarding behavior tested by our classification procedures were unrelated to each other.

There remained, however, the question of whether the behavior of the beneficent subjects confirmed the reformulated versions of Hypotheses 2 and 3. Were the beneficent subjects better at predicting their partner's behavior in the prisoner's dilemma games than were the self-centered ones (Hypothesis 2a)? Moreover, were beneficent subjects less inclined to exploit their partner's expected cooperation (Hypothesis 3a)?

Utilizing the first round data from both DIG 1 and DIG 2, beneficent subjects were not found to be better at identifying a willingness to cooperate than were their self-centered counterparts. Hypothesis 2a, therefore, could not be confirmed. Contrary to our expectations,

---

[23] We also examined whether the type of partner (cooperative or individualistic) played a role in influencing the decisions made by subjects. Only cooperative subjects exhibited a behavioral change (i.e. they reduced decisions for cooperation by 13 percent if they had an individualistic partner). This difference, however, was not significant ($P = 0.122$; binomial test, one-tailed).

[24] In the decomposed games, neither the observed vectors nor the measured consistency were affected by the order of the two classification procedures (vectors: $P = 0.818$, consistency: $P = 0.258$; Mann–Whitney *U*-test).

[25] The six subjects who played another version of the dictator game in the pilot experiment were excluded.

DIG 1 even revealed a significant relationship between the predictions made by self-centered subjects and the subsequent decisions of their partners ($P = 0.007$; Fisher-test).[26]

With regard to Hypothesis 3a, we found support only in DIG 1. That is, we observed only in DIG 1 that beneficent subjects exploited their partner's expected cooperative decision significantly less often than did the self-centered subjects (first round: DIG 1, $P = 0.000$ and DIG 2, $P = 0.043$; binomial test, one-tailed; average exploitation in all three rounds: DIG 1, $P = 0.000$ and DIG 2, $P = 0.074$; Mann–Whitney U-test).

In DIG 2, we did not find support for either Hypothesis 2a or Hypothesis 3a. Even worse, regarding DIG 1, we observed that only the self-centered subjects' predictions were significantly better than chance, while at the same time, these subjects exploited the predicted cooperation of their partners significantly more often.[27]

4. Conclusions

In this paper, we examined whether individuals are able to signal and to recognize the willingness to cooperate. Specifically, we were interested in the question of whether these capabilities are more developed in individuals who are inclined to cooperate themselves. Using the decomposed game technique as a procedure for classifying subjects as cooperative or individualistic, our results revealed that cooperative individuals are somewhat better at predicting their partner's decisions in one-shot prisoner's dilemma games than are the individualistic ones. Our analysis also demonstrated that, after communication, subjects are less inclined to exploit their partner's anticipated cooperation. Interestingly, this held for both cooperative and individualistic subjects. With regard to these findings, we can conclude that the motivation for cooperation, at least in anonymity, is closely related to the capability to signal and to recognize a willingness to cooperate. Individuals that are inclined to cooperate, however, can utilize this capability only if they have the opportunity to communicate. As long as communication is possible, they are able to interact selectively in order to realize the advantages of cooperation, which might compensate for the disadvantages that emerge from their "unconditional" cooperation in anonymity. The findings in this paper underline the importance of communication for individual cooperation and, moreover, support the empirical relevance of evolutionary explanations of cooperation where communication plays an essential role.

Surprisingly, we discovered that fair behavior in the dictator game did not reveal the same characteristics as did cooperative behavior in the decomposed games. Fair behavior also yielded very different results regarding our hypotheses. This gives rise to the conclusion

[26] For all analyses of Hypotheses 2a and 3a, we used the same sample (DIG 1: 35 beneficent and 21 self-centered subjects ($N = 56$), DIG 2: 22 beneficent and 29 self-centered subjects ($N = 51$)). That is, we excluded all subjects from the sample, who personally knew one of their partners in the three rounds or whose partner had explicitly stated an intention to defect in one of the rounds.

[27] Examining whether the type of partner affected the decisions made revealed that beneficent subjects did not reduce the number of decisions for cooperation when their partner was classified as self-centered. This held for both DIG 1 and DIG 2. With regard to DIG 2, beneficent subjects even defected significantly more when their partner was also beneficent ($P = 0.007$; binomial test, one-tailed).

that, contrary to the proposition suggested by Frank (1988), the existence of fair behavior requires a somewhat different explanation than the existence of cooperation.

Acknowledgements

I would like to thank all of the students and colleagues who supported me in conducting my research. I am especially indebted to Professor Joachim Weimann for his encouragement and advice. I have also benefited from the comments made by Professor Werner Gueth as well as those made by my colleagues Dr. Axel Ockenfels and Atanasios Mitropoulos. I gratefully acknowledge the financial support received from the Deutsche Forschungsgemeinschaft.

Appendix A. Decomposed game technique

Using the decomposed game technique it is assumed that the *i*th subject has an utility function $U_i = (x, y)$ which is a linear combination of the own payoff *x* and the other's payoff *y* (Wyer, 1969; Griesinger and Livingston, 1973). Furthermore, the weights m_i in this linear combination are behavioral parameters denoting the *i*th subject's particular "social motivation".

$$U_i = m_{xi}a + m_{yi}b$$

where m_{xi} is the importance given to the own payoff by the *i*th subject, and m_{yi} is the importance given to the other's payoff by the *i*th subject.

In the two-person geometric model developed for this technique by Griesinger and Livingston (1973), the utility function for the *i*th subject, with a certain social motivation, is characterized by a motivational vector (m_{xi}, m_{yi}) of infinite length extending from the origin into a two-dimensional space defined by the own payoff *x* (abscissa) and the other's payoff *y* (ordinate). In particular, a player's social motivation can be represented by the angle θ_i (measured from the abscissa) to the motivational vector determined directly from m_{xi} and m_{yi}.

$$\theta_i \equiv \tan^{-1}\frac{m_{yi}}{m_{xi}}$$

In social psychology, six basic classifications of social motivations are distinguished: altruism ($\theta_{\text{al}} = 90^\circ$), cooperation ($\theta_{\text{co}} = 45^\circ$), individualism ($\theta_{\text{in}} = 0^\circ$), competition ($\theta_{\text{com}} = 315^\circ$), aggression ($\theta_{\text{ag}} = 270^\circ$), sadomasochism ($\theta_{\text{sa}} = 225^\circ$), masochism ($\theta_{\text{mas}} = 180^\circ$), and martyrdom ($\theta_{\text{ma}} = 135^\circ$) (see Griesinger and Livingston, 1973; McClintock and vanAvermaet, 1982) (Fig. 3).

Choosing one of the two own-other payoff combinations *j* in a decomposed game means deciding on an ordered pair of values (x_j, y_j) in the own-other payoff space, or equivalently, on a vector from the origin to the selected ordered pair. The player is assumed to choose that own-other payoff combination that yields the highest level of utility, i.e. that ordered pair whose vector has the greatest projection on the motivational vector characterizing the player's social motivation (Griesinger and Livingston, 1973, p. 177).

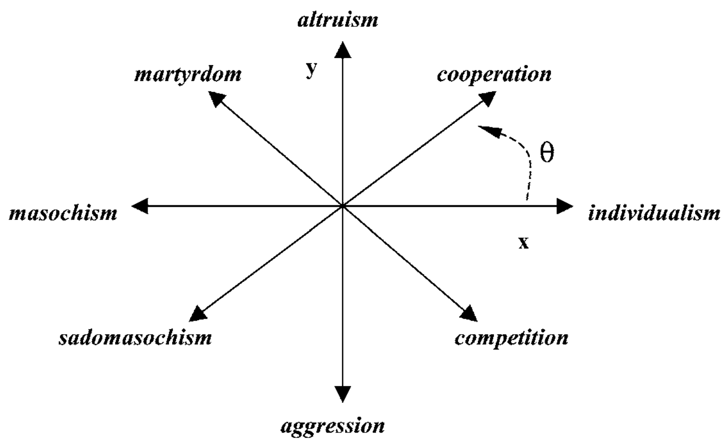

Fig. 3. Vectors defining the basic social motivations.

Determining the social motivation of subjects experimentally is done by confronting them with a series of decomposed games, where they must decide between two own-other payoff combinations. The pairs of payoffs lie equally spaced around a circle. Each pair consists of two adjacent own-other payoff combinations. By adding up the chosen amounts separately, for the subject and the partner, an estimate of the weights (m^*_{xi}, m^*_{yi}) assigned by the subject to the own payoff and to the other's payoff is provided (Liebrand, 1984, p. 250). Using these estimated weights, an estimate of the angle θ_i can be obtained (θ^*_i), and thus, the motivational vector can be estimated:

$$\theta^*_i = \tan^{-1}\frac{m^*_{yi}}{m^*_{xi}}$$

where

$$m^*_{xi} = \sum x_j, \quad m^*_{yi} = \sum y_j$$

In the decomposed game technique employed in this paper, the estimated weights also determine the length of the motivational vector, which provides an index of the consistency of subjects' decisions (M_i):

$$M_i = ((m^*_{xi})^2 + (m^*_{yi})^2)^{1/2}$$

References

Adolphs, R., Damasio, H., Tranel, D., Damasio, A.R., 1996. Cortical systems for the recognition of emotion in facial expressions. The Journal of Neuroscience 7678–7687.

Amann, E., Yang, C.-L., 1998. Sophistication and the persistence of cooperation. Journal of Economic Behavior and Organization 37, 91–105.

Anderson, S.P., Goeree, J.K., Holt, C.A., 1998. A theoretical analysis of altruism and decision errors in public good games. Journal of Public Economics 70, 297–323.

Andreoni, J., 1993. An experimental test of the public goods crowding-out hypothesis. American Economic Review 83, 1317–1327.
Andreoni, J., Miller, J.H., 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. The Economic Journal 103, 570–585.
Andreoni, J., Miller, J.H., 1995. Giving according to GARP: an experimental study of rationality and altruism. Working paper, University of Wisconsin, Madison.
Barkow, J.H., Cosmides, L., Tooby, J., 1992. The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford University Press, New York.
Bohnet, I., Frey, B.S., 1995. Ist Reden Silber und Schweigen Gold? Eine ökonomische Analyse. Zeitschrift für Wirtschafts und Sozialwissenschaften 115, 169–209.
Bolton, G.E., Ockenfels, A., 2000. ERC: A theory of equity, reciprocity, and competition. American Economic Review 90, 166–193.
Brandts, J., Schram, A., 1996. Cooperation and noise in public goods experiments: applying the contribution function approach. Journal of Public Economics 79, 399–427.
Brosig, J., Ockenfels, A., Weimann, J., 2000. The effect of communication media on cooperation. Working paper, Otto-von-Guericke-Universität, Magdeburg.
Buss, D.M., 1999. Evolutionary Psychology: The New Science of the Mind. Allyn & Bacon, Boston.
Cain, M., 1998. An experimental investigation of motives and information in the prisoner's dilemma game. Working paper, University of Mississippi.
Charness, G., 2000. Bargaining efficiency and screening: an experimental investigation. Journal of Economic Behavior and Organization 42, 285–304.
Charness, G., Rabin, M., 2000. Social preferences: some simple tests and a new model. Working paper, University of California, Berkeley.
Cooper, R., DeJong, D.V., Forsythe, R., Ross, T.W., 1996. Cooperation without reputation: experimental evidence from prisoner's dilemma games. Games and Economic Behavior 12, 187–218.
Dawes, R.M., McTavish, J., Shaklee, H., 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. Journal of Personality and Social Psychology 35, 1–11.
Dufwenberg, M., Kirchsteiger, G., 1998. A theory of sequential reciprocity. Center Discussion Paper No. 9837, Tilburg University.
Falk, A., Fischbacher, U., 1998. A Theory of Reciprocity. University of Zurich, Mimeo.
Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114, 817–868.
Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games and Economic Behavior 6, 347–369.
Frank, R.H., 1987. If homo economicus could choose his own utility function, would he want one with a conscience? American Economic Review 77, 593–604.
Frank, R.H., 1988. Passions Within Reason. The Strategic Role of the Emotions. WW Norton & Co., New York.
Frank, R.H., Gilovich, T., Regan, D.T., 1993. The evolution of one-shot cooperation: an experiment. Ethology and Sociobiology 14, 247–256.
Griesinger, D.W., Livingston, J.W., 1973. Toward a model of interpersonal motivation in experimental games. Behavioral Science 18, 173–188.
Guttman, J.M., 1996. Rational actors, tit-for-tat types, and the evolution of cooperation. Journal of Economic Behavior and Organization 29, 27–56.
Hackett, S., Schlager, E., Walker, J., 1994. The role of communication in resolving commons dilemmas: experimental evidence with heterogeneous appropriators. Journal of Environmental Economics and Management 27, 99–126.
Hammerstein, P., Selten, R., 1994. Game theory and evolutionary biology. In: Aumann, R.J., Hart, S. (Eds.), Handbook of Game Theory, Vol. 2. Elsevier, Amsterdam, pp. 929–993.
Harrington, J.E., 1989. If homo economicus could choose his own utility function, would he want one with a conscience? Comment. American Economic Review 79, 588–593.
Isaac, R.M., Walker, J.M., 1988. Communication and free-riding behavior: the voluntary contribution mechanism. Economic Inquiry 26, 585–608.
Isaac, R.M., McCue, K.F., Plott, C., 1985. Public goods provision in an experimental environment. Journal of Public Economics 26, 51–74.

Isaac, R.M., Walker, J.M., Williams, A.W., 1994. Group size and the voluntary provision of public goods. Journal of Public Economics 54, 1–36.

Kahneman, D., Knetsch, J.L., Thaler, R.H., 1986. Fairness and the assumptions of economics. Journal of Business 59, 285–300.

Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoner's dilemma. Journal of Economic Theory 27, 245–252.

Ledyard, J., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–194.

Liebrand, W.G., 1984. The effect of social motives, communication and group size on behavior in an *N*-person multi-stage mixed-motive game. European Journal of Social Psychology 14, 239–264.

McClintock, C.G., vanAvermaet, E., 1982. Social values and rules of fairness: a theoretical perspective. In: Derlega, V.J., Grzelak, J. (Eds.), Cooperation and Helping Behavior. Theories and Research. Academic Press, New York, pp. 43–71.

McClintock, C.G., Liebrand, W.B.G., 1988. Role of interdependence structure, individual value orientation, and another's strategy in social decision making: a transformational analysis. Journal of Personality and Social Psychology 55, 396–409.

McKelvey, R.D., Palfrey, T.R., 1992. An experimental study of the centipede game. Econometrica 60, 803–836.

Ockenfels, A., Selten, R., 2000. An experiment on the hypothesis of involuntary truth-signalling in bargaining. Games and Economic Behavior 33, 90–116.

Offerman, T., Sonnemans, J., Schram, A., 1996. Value orientations, expectations, and voluntary contributions in public goods. The Economic Journal 106, 817–845.

Orbell, J.M., van de Kragt, A.J.C., Dawes, R.M., 1988. Explaining discussion-induced cooperation. Journal of Personality and Social Psychology 54, 811–819.

Ostrom, E., Walker, J.M., 1991. Communication in a commons: cooperation without external enforcement. In: Palfrey, T. (Ed.), Laboratory Research in Political Economy. University of Michigan Press, Ann Arbor, pp. 287–322.

Ostrom, E., Gardner, R., Walker, J., 1994. Rules, Games, and Common Pool Resources. University of Michigan Press, Ann Arbor.

Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in linear public good experiments: how much and why? American Economic Review 87, 829–846.

Rabin, M., 1993. Incorporating fairness into game theory and economics. American Economic Review 83, 1281–1302.

Robson, A.J., 1990. Efficiency in evolutionary games: Darwin, Nash and the secret handshake. Journal of Theoretical Biology 144, 376–396.

Sally, D., 1995. Conversation and cooperation in social dilemmas. A *meta*-analysis of experiments from 1958 to 1992. Rationality and Society 7, 58–92.

Vogt, C., 2000. The evolution of cooperation in prisoner's dilemma with an endogenous learning mutant. Journal of Economic Behavior and Organization 42, 347–373.

Weibull, J.W., 1995. Evolutionary Game Theory. MIT Press, Cambridge.

Weimann, J., 1994. Individual behavior in a free riding experiment. Journal of Public Economics 54, 185–200.

Wyer, R.S., 1969. Prediction of behavior in two-person games. Journal of Personality and Social Psychology 13, 222–238.

Yang, C.-L., 1993. Evolution of cooperation in prisoner's dilemma game with a knowledgeable mutant. Discussion paper No. 93-09, Universität Dortmund.