

Social Framing Effects: Preferences or Beliefs?

Tore Ellingsen
Stockholm School of Economics

Magnus Johannesson
Stockholm School of Economics

Johanna Mollerstrom
Harvard University

Sara Munkhammar

This version: 19 January 2011

Abstract

In an otherwise neutrally described Prisoners' dilemma experiment, we document that behavior is more likely to be cooperative when the game is called the Community Game than when it is called the Stock Market Game. However, the difference vanishes when only one of the subjects is in control of their own action. The social framing effect also vanishes when the game is played sequentially. These findings are inconsistent with the hypothesis that people's desires are affected by social frames. Instead, they suggest that social frames are coordination devices. That is, social frames enter people's beliefs rather than their preferences.

Acknowledgement: Thanks to James Andreoni, Roland Bénabou, Colin Camerer, David Cesarini, Vincent Crawford, Dirk Engelmann, Ernst Fehr, Michael Kosfeld, John List, Erik Mohlin, Martin Nowak, Charles Plott, Jean Tirole, Björn Wallace, Robert Östling, and many seminar participants for helpful comments and discussions, and to Josefin Grankvist, Ola Granström, Eva Hildén Smith, Erik Höglin, Erica Jernelöv, Erik Lindqvist, Tobias Lundquist, Erik Mohlin, Per Sonnerby, Eva Ranehill, Björn Tyrefors, Åsa Törlén, Niklas Zethraeus and Eva Bring for research assistance. Financial support from the Torsten and Ragnar Söderberg Foundation, the Jan Wallander and Tom Hedelius Foundation and the Swedish Research Council is gratefully acknowledged.

1. INTRODUCTION

In a seminal experiment, Deutsch (1958) showed that behavior in a Prisoners' dilemma depends on whether experimental subjects are induced to feel cooperative or individualistic before making their choice. While Deutsch's instructions were heavily loaded,¹ it has long been clear that subtler contextual manipulations may affect behavior too. Eiser and Bhavnani (1974) find that behavior in a Prisoners' dilemma is more cooperative when the situation is framed as an international negotiation than when it is framed as a business transaction. Likewise, subjects cooperate more in a "social exchange study" than in a "business transaction study" (Batson and Moran, 1999), and substantially more in a "community game" than in a "Wall Street game" (Kay and Ross, 2003; Liberman, Samuels, and Ross, 2004), even when the subjects' instructions are otherwise neutral.²

Such context sensitivity has been interpreted as bad news for utility theory in general (Weber, Kopelman, and Messick, 2004) and for social preference theories in particular (Levitt and List, 2007). To the extent that people can be seen as maximizing utility at all, it appears that the utility function must include situational elements that conventional theory leaves out. At a more practical level, the results have been used to criticize economists' emphasis on material incentives. By triggering a selfish social frame, material incentives could potentially reduce, for example, employee effort (Frey and Osterloh, 2005; Pfeffer, 2007), legal compliance (Tyran and Feld, 2006; Bohnet and Cooter, undated), and other prosocial

¹ According to Deutsch (1960), in the cooperative condition, the beginning of the instruction was: "Before you start playing the game, let me emphasize that in playing the game you should consider yourself to be partners. You're interested in your partner's welfare as well as in your own." In the individualistic condition, the beginning of the instruction instead was: "Before you start playing the game, let me emphasize that in playing the game your *only* motivation should be to win as much money as you can for yourself. You have no interest whatsoever in whether the other person wins or loses or in how much he wins or loses."

² From now on, we use the term social frame in the narrow sense of "the name of the game" – the labeling of the situation. Other studies that investigate the impact of game labels or strategy labels in social dilemmas include, *inter alia*, Andreoni (1995), Brewer and Kramer (1986), Cookson (2000), Pillutla and Chen (1999), Rege and Telle (2004), van Dijk and Wilke (2000), and Zhong, Loewenstein, and Murnighan (2007). Labels have also been shown to affect cooperative behavior in other games; see for example Larrick and Blount (1997). Whereas we are only concerned with the effect of labeling on cooperation in social situations, Tversky and Kahneman (1981) (who coined the "framing effect" concept) showed that wording can have a significant impact on individual choice as well; see Levin, Schneider, and Gaeth (1998) for a survey of individual choice effects of wording.

behaviors (Koneberg, Yaish, and Stocké, 2010). Through this channel, the very language and assumptions of economics could be eroding cooperation (Ferraro, Pfeffer, and Sutton, 2005).

However, the lessons from the experimental findings are less obvious than they may first appear. The social framing results have several possible explanations, and additional evidence is needed to discriminate between them. As Camerer (2003, p75) puts it: “There is little doubt that describing games differently can affect behavior; the key step is figuring out what *general* principles (or theory of framing) can be abstracted from labeling effects.” Our purpose here is to provide new evidence that helps to elucidate these general principles.

It is possible to distinguish three broad classes of social framing theories. Perhaps the most widely held theory is that people have a built-in tendency to do what the situation asks of them (Montgomery, 1998; Weber, Kopelman, and Messick, 2004). Since the explanation is based on the “logic of appropriateness,” as articulated by March (1994, Chapter 2), we refer to it as the *appropriateness hypothesis*. A specific version of the appropriateness hypothesis is that the decision frame affects people’s propensity to engage in team reasoning, defined as choosing the action that is best for the group rather than that which is in the own best interest; see Sugden (1993) and Bacharach (1999).

A separate but closely related hypothesis is that people respond to social frames because the frame affects how others interpret their behavior, which in turn determines their social esteem. This is the *social image hypothesis*.³

A third class of theories of social framing effects is that people care not only about their own material payoffs, but also about others’ actions (Sen, 1967), intentions (Rabin, 1993) or material payoffs (Becker, 1974; Fehr and Schmidt, 1999), in which case a Prisoners’ dilemma in material payoffs may be transformed into a “Stag hunt” in utilities. More precisely, the *game form* (which summarizes the objective features of strategies and payoffs) is a Prisoners’ dilemma, but the *game* (which involves von Neumann-Morgenstern utilities) is Stag hunt.⁴ Since a Stag hunt game has two pure strategy Nash equilibria, the frame can be used as an equilibrium selection device, as noted by Rabin (1998) and Fehr and Schmidt (2006). At the

³ There is a sizeable literature documenting that people are motivated by social esteem considerations; see for example Brennan and Pettit (2004), Ellingsen and Johannesson (2007), Andreoni and Bernheim (2009) and the references therein.

⁴ The term “game form” originates from Gibbard (1973, p587), and is used synonymously with “mechanism” in the mechanism design literature. It is a description of the link between strategies and outcomes. Utility functions, in turn, link outcomes to real numbers (utilities).

outset, we lump all these social preference models together and refer to them as the *coordination hypothesis*.

In order to evaluate the relevant importance of the three hypotheses, we first state them formally. We then conduct three separate one-shot Prisoners' dilemma experiments, in which we systematically vary features of the game. Our most striking finding is that there is indeed a significant social framing effect when subjects make their decisions simultaneously, as in our first experiment, but not when decisions are made sequentially, as in our third experiment. This finding is consistent with the coordination hypothesis, but inconsistent with the appropriateness hypothesis. Briefly, the argument runs as follows: Because a sequential Stag hunt game has a unique subgame perfect equilibrium, there is no room for a coordinating role of labels. On the other hand, the logic of appropriateness should affect behavior regardless of whether moves are simultaneous or sequential. Our findings thus suggest that framing effects in social dilemmas are well explained within the rational choice paradigm, without any appeal to context-dependent preferences.

The paper is most closely related to Liberman, Samuels, and Ross (2004), whose results were circulated already in the early 1990's. They report on three studies. The first study compares behavior in a seven-round Prisoners' dilemma under a "Wall Street Game" frame to the corresponding behavior under a "Community Game" frame, using a selected group of 48 male college students.⁵ The second study instead uses 40 Israeli pilot trainees, and the labels Bursa Game and Kommuna Game, but is otherwise similar. In both studies, cooperation rates are significantly higher under the Community/Kommuna Game frame. In the second study, both the pilot trainees themselves and their (flight) instructors are asked to make predictions about others' first-round behavior. On average, the participants are more optimistic regarding others' cooperation in the Kommuna Game than in the Bursa Game, but no such difference is observed among instructors. Moreover, participants expecting first-round cooperation are relatively likely to cooperate in the Kommuna Game, but not in the Bursa Game. Finally, in the third study, college students who had not participated in Study 1 were asked to predict first-round choices. Like the flight instructors in Study 2, these subjects failed to predict the large difference in cooperation rates between the two frames, suggesting that beliefs depend on whether one is a participant in the situation or not.

⁵ One purpose of the study was to compare the influence of the social context with that of presumed personality characteristics. The subjects had been chosen by peers based on their likely propensity to cooperate.

Besides involving a much larger number of subjects, and hence having more statistical power, our experiments provide qualitatively new insights. First, by considering a one-shot Prisoners' dilemma, we narrow down the set of explanations: We rule out the possible objection that even selfish materialists could find it in their interest to cooperate in the first round of a finitely repeated game, either because of uncertainty about the opponent's type (Kreps et al, 1982) or because the payoff loss from one round's cooperation is small enough to neglect (Radner, 1986). Second, and more importantly, we use variation in available strategies and information to discriminate between different explanations for social framing effects. Liberman, Samuels, and Ross (2004) cannot rule out the possibility that the frame's primary effect is on the preferences, and that beliefs only change as a result of the preference change. Our evidence suggests that the social frame only affects behavior through the beliefs, and not through preferences.

Other closely related experimental studies include Cookson (2000) and Rege and Telle (2004), who study social framing effects in public goods contribution games. Both these papers find that a more community oriented frame creates more cooperation, although in the latter study the effect is only marginally statistically significant, probably due to a small number of subjects per treatment.⁶

We are not the first to utilize a sequential Prisoners' dilemma to disentangle preferences and beliefs. In a study of in-group favoritism Yamagishi and Kiyonari (2000) find that there is more in-group favoritism in Prisoners' dilemmas with simultaneous play than in games with sequential play.⁷ Although Yamagishi and Kiyonari do not explicitly invoke the game theoretic argument, it is clear that their idea is similar to ours: The sharp reduction of in-group favoritism in the sequential setting suggests that the in-group favoritism in the simultaneous setting is driven primarily by expectations, not preferences. However, Yamagishi and Kiyonari only study the behavior of first-movers, whereas our strongest evidence comes from the absence of a framing effect among second movers in the sequential game.

⁶ Note however that two recent studies, Brandts and Schwiieren (2009) and Dufwenberg, Gächter and Hennig-Schmidt (2010), report similar experiments that fail to establish the expected framing effect. In the latter, it turns out that there is a natural explanation – the particular subjects' negative view of their own community – but in the former there is no obvious reason for the lack of a framing effect.

⁷ In-group favoritism refers to the phenomenon that people behave more favorably toward members of the own group than toward non-members. We refer to Chen and Li (2009) for an extensive review and experimental evaluation of in-group favoritism.

Our findings also relate quite closely to Bohnet and Cooter (undated), who experimentally compare the behavioral impact of small penalties across different game forms. They find little effect of penalties in a many-player Prisoners' dilemma, but large effects in a coordination game. A natural interpretation is that the small penalties for defection from socially optimal actions moved the beliefs in a favorable direction, and that such movement only matters in a coordination game. However, as we point out, one cannot from looking at material payoffs alone infer what the real game is; for two conditional cooperators, the Prisoners dilemma game form is a coordination game. Thus, the above natural interpretation requires an independent argument for why material payoffs and utilities are likely to coincide in this case.

The paper is organized as follows. Section 2 briefly discusses the different theories. Section 3 describes our first study, which establishes both that social framing effects exist and that they can be removed by suitable manipulations of the environment. More precisely, the study shows that the framing effect vanishes when the opponent is unaware of what game is being played and the opponent's action is controlled by an appropriately programmed computer. The second study, reported in Section 4, shows that the framing effect remains absent under otherwise similar circumstances even if the opponent is informed, a finding which goes against the notion that people cooperate in the Community Game in order to impress their opponent with their altruism. The third and final study, reported in Section 5, shows that there is no framing effect in the sequential Prisoners' dilemma. Section 6 concludes.

2. THEORY

Consider two players facing the actions and material payoffs (the game form) depicted in Figure 1.

	C	D
C	c,c	s,w
D	w,s	d,d

Figure 1: Material payoffs - the game form

Let $w > c > d > s$. Moreover, let $s + w < 2c$. Thus, the sum of the material payoffs is largest if both players choose action C. However, if the players are selfish materialists, they will both be

playing D, since D maximizes the own material payoff regardless of what the opponent does. That is, the game form in Figure 1 is a Prisoners' dilemma. (The action labels are chosen to indicate Cooperation and Defection, respectively, and the letters for the payoff parameters c and d are chosen accordingly. In line with the classical treatments of the Prisoners' dilemma, the letters s and w indicate the “sucker” payoff and the “winner” payoff respectively.

From now on, we say that a player is selfish if she cares only about the own material payoff. All other player types are considered to be unselfish to some degree.

We introduce the following pieces of notation. Let $S_i \in \{C, D\}$ be a pure action for player i and let $m_i(S_i, S_j) \in \mathbb{R}$ be player i 's material payoff given the strategy profile (S_i, S_j) . Players may care about the opponent's payoff too, that is, they may have social preferences. Players may also care about non-material outcomes, such as the opponent's belief about their social preferences. To capture such preferences, let \mathcal{T} be a finite set of possible player types, with τ as a typical element; let τ_{ij} denote player i 's expectation about player j 's type; and let τ_{ji} denote player i 's expectation about τ_{ji} .⁸ Finally, let \mathcal{F} denote the set of frames, with typical element F . In general, player i 's utility can thus be written as a function $U_i(m_i, m_j, \tau, \tau_{ij}, \tau_{ji}, F)$.

Let us now consider some specific examples of how the frame may enter players' preferences and beliefs respectively.

2.1 Frame-dependent preferences

The frame could be affecting the degree of unselfishness. We consider three examples.

Suppose first that players are altruists, but that their altruism depends on the frame F . Specifically, let each player i assign utility $\alpha(F)m_j$ to the opponent's material payoff m_j . That is, utility can be written

$$U_i = m_i + \alpha(F)m_j. \quad (1)$$

⁸ For early formal models of social esteem concerns, see Bernheim (1994), Ireland (1994) and Glazer and Konrad (1996). For recent extensions and applications, see Bénabou and Tirole (2006) and Ellingsen and Johannesson (2008). Experimental evidence suggests that people care about what others think about their actions even if the interaction itself is anonymous; see Dana, Cain, and Dawes (2006), Broberg, Ellingsen and Johannesson (2007), and Lazear, Malmendier, and Weber (2010). Beliefs may affect preferences also in a model without “types” – see Dufwenberg, Gächter, and Hennig Schmidt (2010) for references and an application to framing effects.

Suppose these preferences are common knowledge. The game corresponding to the game form in Figure 1 is then given by the bi-matrix depicted in Figure 1a.

	C	D
C	$c+\alpha(F)c, c+\alpha(F)c$	$s+\alpha(F)w, w+\alpha(F)s$
D	$w+\alpha(F)s, s+\alpha(F)w$	$d+\alpha(F)d, d+\alpha(F)d$

Figure 1a: The game played by frame-sensitive altruists

If players are sufficiently altruistic, the game is not a Prisoners' dilemma. For example, if $\alpha > \max\{(w-c)/(c-s), (d-s)/(w-d)\}$, each player's dominant choice is to play the cooperative action C, whereas action D remains dominant if $\alpha < \min\{(w-c)/(c-s), (d-s)/(w-d)\}$. (For intermediate values of α , the game is Stag hunt if $(d-s)/(w-d) > (w-c)/(c-s)$, and Chicken if $(d-s)/(w-d) < (w-c)/(c-s)$.) Note that team reasoning, as defined by Sugden (1993), can be seen as the special case in which $\alpha=1$, implying that the cooperative action A is dominant.

A natural interpretation of the players' altruistic concern $\alpha(F)$ is that it captures an internalized efficiency norm.⁹ If this efficiency norm is strong enough, it suffices to sustain cooperation. Observe that beliefs about the opponent are irrelevant whenever a player has a dominant action. Thus, even if players are heterogeneous, differing with respect to their propensity to comply with norms, the belief about the opponent's propensity does not enter into the decision problem.

If people obtain more utility from obeying the efficiency norm when the game form is called a "Community Game" than when it is called a "Stock Market Game," the propensity to play A will tend to be higher in the former case.

The social esteem hypothesis says that people are concerned about what others may think about them. For example, a player may get positive utility from believing that the opponent believes that she is altruistic (or obeys the efficiency norm). Formally, player i 's belief about player j 's belief will then enter player i 's utility function. For example, suppose that each player is either selfish ($\tau=0$) or altruistic ($\tau=\alpha(F)$), and that players' desire for social esteem is independent of their actual altruism, but possibly dependent on the frame. Then player i 's utility function can be written

$$U_i = m_i + \alpha(F)m_j + v(\alpha_{ji}, F). \quad (2)$$

⁹ See, e.g., Andreoni and Bernheim (2009) or Krupka and Weber (2009) for richer models of norm compliance based on internalization.

For simplicity, assume that $v(0,F)=0 < v(a,F)$ for all $a > 0$ and any frame F – that is, selfishness is never a source of esteem.

With these preferences, the Prisoners’ dilemma turns into a (two-sided) signaling problem, in which players may cooperate not only because they are altruistic, but also in order to *convey the impression* that they are altruistic. If the value of looking altruistic, $v(\alpha_{ji},F)$, is greater under one frame than another, then this hypothesis works in essentially the same way as the appropriateness hypothesis.¹⁰ However, since only the social esteem considerations are affected by external observability it is still possible to distinguish between the two hypotheses.

2.2 Frame-dependent beliefs

Let us next consider models in which frames do not enter preferences, but may be entering beliefs instead. If the game has multiple equilibria, the frame may then affect equilibrium selection. There are a variety of social preferences that transform a Prisoners’ dilemma game form into a game with multiple equilibria.

Suppose first that players desire to behave altruistically if and only if their material payoff is no smaller than that of the opponent. That is, their utility function takes the form

$$U_i = m_i + \alpha I m_j, \tag{3}$$

where I is an indicator variable taking the value 1 if $m_i \geq m_j$ and 0 otherwise.¹¹ Then, if preferences are common knowledge, the game is as depicted in Figure 1b.

	C	D
C	$c+\alpha c, c+\alpha c$	$s, w+\alpha s$
D	$w+\alpha s, s$	d, d

Figure 1b: The game with altruism towards a worse off opponent

¹⁰ More formally, there will be an open set of parameters such that, in the unique perfect Bayesian equilibrium that satisfies the Intuitive Criterion, the altruists play C under one frame, but D under the other. (Egoists play D under either frame.)

¹¹ The key to our analysis is that altruism is greater when players are ahead, not that it vanishes completely when they are behind. However, this formulation is particularly simple. See Charness and Rabin, 2002, for a detailed discussion of the relevance of such social welfare preferences.

If $\alpha > (w-c)/(c-s)$, the game has two pure strategy equilibria, namely (C,C) and (D,D). The former equilibrium Pareto-dominates the latter. If, in addition, s is sufficiently small, (D,D) is risk-dominant. That is, the game is not Prisoners' dilemma, but Stag hunt.

To the best of our knowledge, this kind of argument was first made by Sen (1967). However, Sen invoked the concept of "conditional cooperation," which is a general description of preferences over alternative actions in the specific situation of a social dilemma rather than a specific preference ordering over a general set of outcomes. As is well known, there are many other general preference orderings over outcomes that may also give rise to "conditional cooperation" in social dilemmas.

For example, as noted by Fehr and Schmidt (2006), an analogous argument holds if both players dislike "taking advantage of" their opponent. (Many people dislike even more to be taken advantage of, but for the current argument it is only the aversion to advantageous inequality that matters.) Specifically, suppose that players' utility can be written

$$U_i = m_i - \beta \max\{0, m_i - m_j\}. \quad (4)$$

Then, if the utilities are common knowledge, the corresponding game is depicted in Figure 1c.

	C	D
C	c, c	$s, w - \beta(w - s)$
D	$w - \beta(w - s), s$	d, d

Figure 1c: The game when players are averse to advantageous inequality

If $\beta > (w-c)/(c-s)$, the game is Stag hunt, with pure strategy equilibria (C,C) and (D,D).

There are several other variants of the above argument, including the conditional fairness model of López-Pérez (2008) as well as the intention-based fairness model of Rabin (1993). Taking inspiration from the theory and findings of Charness and Rabin (2002), here is one model that we like particularly well, and which will also turn out to rationalize all the data. Suppose there are two types of players, egoists and conditional altruists with the utility function

$$U_i = m_i + \alpha I \alpha_{ij} m_j, \quad (5)$$

where I is an indicator variable taking the value 1 if $m_i \geq m_j$ and 0 otherwise. That is, altruism is only triggered when the opponent is behind and is not believed to be an egoist. Then, if α_{ij} and α are sufficiently large, the resulting incomplete information game has two pure strategy

Bayesian Nash equilibria – one in which conditional altruists play C and one in which they play D. Egoists, of course, always play D.

When the game has multiple equilibria, as in models (2)-(5), it is a short step to realize that the frame can be used as a coordination device, or focal point (Schelling, 1960). Moreover, in order for the frame to have an effect on the agents' behavior, it is not necessary to assume that preferences change. Instead, since the social frame only affects beliefs, this approach is fully compatible with the view that models of preferences ought to be parsimonious and portable across games.¹²

The notion that frames affect coordination is more than a theoretical possibility. There is substantial evidence that people use action labels for coordination purposes; see Mehta, Starmer, and Sugden (1994) and Crawford, Gneezy, and Rottenstreich (2008). A formal model of how this may happen has been developed by Bacharach (1993) and refined through the notion of level-k reasoning by Bacharach and Stahl (2000); see also Bacharach and Bernasconi (1997) and Bardsley et al (2010).¹³

Such a level-k model can also be used to rationalize the impact of the game label, as opposed to action labels, on behavior. Roughly, if player *i* thinks that player *j* (if not selfish), is attracted towards joint payoff maximizing actions under the Community label but towards private payoff maximizing actions under the Stock Market label, then player *i*, if conditionally cooperative for either of the reasons specified above, may also choose to cooperate under the Community label but not under the Stock Market label. – Of course, in order to pursue this logic fully, one should not be considering the Nash equilibria at all, but apply this structural non-equilibrium solution concept throughout.

3. THE FIRST STUDY: PRESENCE AND ABSENCE OF FRAMING EFFECTS

The first experiment was conducted at Södertörn University College and Stockholm School of Economics, both in Stockholm, Sweden, on three different occasions. The first sessions were run at Södertörn in April 2006. Subsequent sessions were run at Södertörn in

¹² For a discussion of the trade-off between fit and parsimony in the modeling of people's preferences, see Sobel (2005).

¹³ Cachon and Camerer (1996) and Rydval and Ortman (2005) demonstrate that loss aversion furnishes another, and possibly related, coordination principle.

November 2006 and Stockholm School of Economics in September 2007. On each occasion the subjects were randomly allocated between four treatments.

In total 448 subjects participated as decision-makers in the experiment. All were freshmen enrolled in a basic microeconomics course. In addition, 220 student subjects participated as recipients in the asymmetric information treatment (described below).

Two of the treatments are intended to investigate whether we can replicate previous findings of social framing effects within an experiment that satisfies current requirements in behavioral economics. Specifically, the sample size is large, real money is at stake and each subject is exposed only to one decision frame. Moreover the social framing is quite light; the name of the game differs across treatments, but otherwise the description of the situation is neutral.

The other two treatments are designed to test whether it is possible to reduce or eliminate any framing effects, by manipulating several features of the situation. This is described in detail below.

3.1 Design

In treatments 1 and 2, henceforth called the symmetric information treatments, subjects are seated in four different rooms. Each subject is, anonymously and randomly, paired with a subject in another room, and both subjects receive identical oral and written instructions. Indeed, with the exception of the name of the game, all subjects in treatments 1 and 2 receive identically worded instructions. In one pair of rooms, the situation is called the Stock Market Game (treatment 1); in the other pair of rooms, it is called the Community Game (treatment 2).

The paired subjects simultaneously choose between two options, denoted A and B respectively. If both subjects choose option A, each earns 50 SEK¹⁴ (Swedish Kronor; \$1≈SEK 7.50 at the time of the experiment). If both subjects choose option B, each earns SEK 20. If one subject chooses option A and the other subject chooses option B, the former earns SEK 5 and the latter earns SEK 80. The associated game form is depicted in Figure 2.

¹⁴ In April 2006, USD 1= SEK 7.6. At the time of the following experiments the krona's exchange rate is slightly better, with the krona hitting its highest value against the dollar (USD 1= SEK 6.7) in September 2008 and September 2009.

	A	B
A	50,50	5,80
B	80,5	20,20

Figure 2: The game form in the experiment

Since each subject earns more by choosing B than by choosing A, the pair of actions (B,B) yields lower payoffs for both subjects than the pair (A,A), and the total payoff from (A,B) or (B,A) is lower than from (A,A) the situation is a true Prisoners' dilemma. Indeed, the game form is a special case of that in Figure 1, but with the letters A and B replacing C and D. In all experiments, we used letters A and B in order to minimize the risk that any of our subjects would associate the labels with particular meanings, such as cooperation and defection. However, for ease of reference, we now revert to using letters C and D.

In treatments 3 and 4, henceforth called the asymmetric information treatments, only one person in each pair was in control of the own decision. These decision-making subjects were given oral and written instructions that differed from the ones given in treatments 1 and 2 only with respect to the matched subject's choice. The matched subject was explained to be an uninformed receiver, whose action is chosen by a computer. The computer would make the opponent's action choices with the same frequencies as actual play in the corresponding active opponent treatment. Only information regarding the procedure was given to the decision-maker, and not the actual frequency. The instructions for treatments 3 and 4 were identical except for the name of the game, which was the Stock Market Game in treatment 3 and the Community Game in treatment 4. The receivers were given written information that they were taking part in an economic experiment, but received no information about why they received a specific payoff.

After the experiment, the participants received information about their matched subject's action and were paid accordingly. Appendices I and II contain translations of the complete experimental instructions.

3.2 Predictions

Let models be indexed by the equation number of the corresponding utility function. All the models reviewed above allow the outcome that there is more cooperation in treatment 2 (standard Community Game) than in treatment 1 (standard Stock Market Game). However, they differ substantially regarding their predictions regarding treatments 3 and 4. If we take the game label to be the frame, and under the assumption that beliefs are really the same in T3

(T4) as in T1 (T2), then Model (1) predicts that the framing effect is the same when comparing T3 and T4 as when comparing T1 and T2.

Model (3) and Model (4), predict that behavior in T3 (T4) is identical to that in T1 (T2) under the sole condition that beliefs are the same in T3 as in T1 and in T4 as in T2. According to these two models it only matters what the opponent ends up doing, not what she knows or wants. Thus, subjects should disregard whether their opponent is active or passive, informed or uninformed.

Model (2) allows behavior to depend on the opponent's information. If anything, there should be more cooperation in T1 than in T3, as social esteem is only at stake in the former case.

Model (5) allows an active subject's behavior to depend on the opponent's freedom of choice.¹⁵ To see how, let μ be the common prior that the opponent is a conditional altruist. Suppose that conditional altruists attempt to coordinate on cooperation in T2 but not T1. As an egoist always defects (plays D), the requirement for a conditional altruist to be playing C in any equilibrium is

$$\mu(50+50\alpha)+(1-\mu)5 \geq \mu(80+5\alpha)+ (1-\mu)20,$$

which simplifies to

$$\mu(3\alpha-1) \geq 1.$$

Since μ must lie between 0 and 1, this in turn boils down to the requirements that $3\alpha-1 \geq 1$, or equivalently $\alpha \geq 2/3$, and $\mu \geq 1/(3\alpha-1)$.

But suppose now instead that one of the players is unable to choose freely – instead the cooperation rate is simply fixed at μ for both egoists and conditional altruists, as will be the case in T4 under our assumption about play in T2. Then, the active player chooses to play C if and only if

$$\mu(50+50\alpha\mu)+(1-\mu)5 \geq \mu(80+5\alpha\mu)+ (1-\mu)(20+20\alpha\mu),$$

¹⁵ McCabe, Rigdon and Smith (2003) compare behavior in a standard trust game with that in an “involuntary” trust game, in which the trustor has no choice but to trust. The trustee is more likely to reward voluntary trust than involuntary trust. Since the prisoners' dilemma is essentially a simultaneous move version of the trust game, their finding is quite relevant here.

which simplifies to

$$\mu((13\mu-4)\alpha-3) \geq 3.$$

Again, we can use the fact that μ lies between 0 and 1 to deduce the necessary condition $\alpha \geq 2/3$, just as before. However, the necessary condition on μ when $\alpha > 2/3$ is stronger than before (as $(13\mu-4)/3 < 1$ for all $\mu < 1$). For example, when $\alpha = 1$, the critical value of μ jumps from $1/2$ in the active opponent case to more than $4/5$ in the passive opponent case. Likewise, if $\alpha = 3$, the critical value of μ jumps from $1/8$ in the active opponent case to more than $1/2$ in the passive opponent case. Thus, Model (5) suggests that a cooperative equilibrium which exists when both players are active could well be vanishing when one player is passive. Intuitively, this happens both because the active player's utility from the (C,C) outcome is smaller under T4 than T2 – as some of the cooperation benefits in T4 go to an egoistic opponent – and because the active player's utility from (D,D) is larger under T4 than T2 – as some defecting opponents are conditional altruists (any defecting active opponent is an egoist).

3.3 Findings

The findings are displayed in Figure 3.¹⁶ They reveal a social framing effect in the symmetric information condition. The fraction of subjects making the cooperative choice increases from 26.4% with the Stock Market Game frame to 44.9% with the Community Game frame. This difference is statistically significant (one-sided z-test; $z=2.972$, $p=0.0015$),¹⁷ rejecting the null hypothesis of zero framing effect.

FIGURE 3 ABOUT HERE

Under asymmetric information, on the other hand, the framing effect is insignificant with 26.8% making the cooperative choice with the Stock Market game frame and 28.7%

¹⁶ Throughout, the confidence intervals are normal approximation intervals; given that our samples are large and that the cooperation probability is not too close to 0 or 1, the normal approximation to the binomial distribution is known to be good. A possible exception is the last pair of staples of Figure 7.

¹⁷ As our alternative hypothesis is one-sided (more cooperation under the Community label), we use one-sided tests for all comparisons of cooperation levels.

with the Community Game frame ($z=0.316$, $p=0.376$). The difference-in-difference between the two information conditions is also statistically significant ($z=1.910$, $p=0.028$), rejecting the null hypothesis that the framing effect is the same under asymmetric information as under symmetric information, and favoring the alternative hypothesis that the framing effect is larger under symmetric information.¹⁸

The finding of a social framing effect in treatments 1 and 2 shows that prior findings are robust to such features as monetary incentives and lightly loaded instructions. It is somewhat less clear how we should interpret the absence of a social framing effect in treatments 3 and 4. Essentially subjects in both treatments 3 and 4 have the same cooperation rates as in treatment 1.

Of the five models presented above, only Model (2) and Model (5) are directly consistent with the evidence.

As mentioned above, Model (1), the appropriateness hypothesis, could be rescued by invoking the argument that the frame is more than just the label. However, Models (3) and (4) can only be rescued by assuming that subjects, erroneously, hold other beliefs in the active condition than in the passive condition. Liberman, Samuels and Ross (2004) elicited beliefs both by participants and non-participants in their social framing experiment, finding that participants' beliefs respond more strongly to the frame than beliefs of non-participants. Similar differences could in principle arise between the active and passive opponent conditions. Arguably, belief elicitation would have enabled us to see whether beliefs vary across conditions. However, belief data have problems of their own. Belief elicitation before subjects choose their action may affect behavior (Croson, 1999), belief elicitation after the action choice may be biased by the chosen action (Dawes, McTavish, and Shaklee, 1977), and in any case the elicited beliefs may be quite different from the subjective beliefs that would rationalize the observed choice (Costa-Gomes and Weizsäcker, 2008). Thus, we think that a better way to control for differences in beliefs would be to induce beliefs directly. For example, one might re-run our four treatments with the modification that all subjects in treatments 1 and 3 (2 and 4) are informed about empirical frequencies in our treatment 1 (2).¹⁹

¹⁸ Whereas the framing effect was stable between our sample at Södertörn and the one at SSE, the cooperation ratios (levels) were not the same. At Södertörn, with a total sample size of 230, the cooperation ratio was 0.327 in treatment 1 and 0.530 in treatment 2 ($z=2.245$, $p=0.012$), and 0.328 in treatment 3 and 0.370 in treatment 4 ($z=0.530$, $p=0.298$). At SSE we had a total sample of 218 and the cooperation ratios were 0.207 in treatment 1 and 0.346 in treatment 2 ($z=1.625$, $p=0.052$), and 0.204 in treatment 3 and 0.204 in treatment 4 ($z=0$, $p=0.5$).

¹⁹ We are grateful to Ernst Fehr for this suggestion.

The subsequent experiments reported below instead pursue different directions, attempting to discriminate between the various hypotheses without eliciting or inducing beliefs.

4. THE SECOND STUDY: SOCIAL ESTEEM?

Our own initial hypothesis was that social framing effects are caused largely by people's desire to look good in the eyes of others, as in Model (2). The findings of the first study are consistent with this hypothesis. The second study is designed to provide a sharper test.

We now modify treatments 3 and 4 of the first study in one crucial respect, namely, by letting the passive player observe the game and the choice of the active player. Call the two new treatments 1' (Stock Market Game) and 2' (Community Game) respectively. If it is the difference in information that created the discrepancy between treatments 2 and 4, then there should now be a similar gap between treatments 1' and 2' as between treatments 1 and 2. To be precise, there should be a similar gap if subjects' expected utility of being esteemed by the opponent is independent of whether subjects learn about the opponent's type. On the other hand, if subjects anticipate that they will experience stronger feelings of pride or shame if they learn what the opponent chooses, framing might matter more in treatments 1 and 2 than in 1' and 2'.

Due to a limited number of available subjects, we refrain from re-running previous treatments, thereby increasing the statistical power for the test between the new treatments 1' and 2'.

In total, 137 subjects participated as decision-makers in the experiment, which was conducted in September 2008. All were freshmen enrolled in a basic microeconomics course at Stockholm School of Economics (SSE). In addition, 137 student subjects participated as passive players. Notice that the subjects have very similar characteristics to the SSE subjects in Study 1. In both cases virtually the entire cohort participates, as participation was the default option for participants in the course. The only difference is that subjects in Study 2 belong to a later cohort. Both experiments were conducted on freshmen very early in the term, and since the program has extremely competitive entry requirements, the pool of students always comprises the top echelon of Swedish students. Since we did not use students from Södertörn (Sn) this time around, we also checked for differences in effects between the two populations in the first study. The relative magnitude of the framing effects in Study 1 is as large at SSE as at

Sn, as discussed in footnote 18, but the baseline level of cooperation is lower. Therefore, any difference in framing effects is unlikely to be caused by subject pool effects.

Figure 4 displays the findings.

FIGURE 4 ABOUT HERE

The difference in behavior across treatments is insignificant ($z=-0.271$, $p=0.393$), and the point estimate has the wrong sign. Therefore, subject to the caveat that feelings of pride and shame could be stronger in the active opponent condition, we reject the hypothesis that the social framing effect in Study 1 was caused by social esteem considerations. That is, we reject Model (2).

5. THE THIRD STUDY: PREFERENCES OR COORDINATION?

Only after conducting the first two studies did it occur to us that there is another straightforward way to distinguish between the appropriateness hypothesis and the coordination hypothesis, namely by letting the moves be sequential instead of simultaneous.²⁰ If a Stag hunt game is played sequentially, the second mover can always assure herself of playing a best reply, and hence the efficient equilibrium is the unique subgame perfect outcome. Thus, if any version of the coordination hypothesis is correct, there should be no social framing effect.²¹ On the other hand, if the appropriateness hypothesis is correct, there ought to be a social framing effect even in the sequential game. In particular, the second mover should be more willing to respond to C by playing C in the Community Game than in the Stock Market Game.

To investigate this issue, we conducted a third studies with two treatments that we call 1'' (Stock Market Game) and 2'' (Community Game). These are similar to treatments 1 and

²⁰ For a detailed study of behavior in sequential Prisoners' dilemmas, see Clark and Sefton (2001).

²¹ Observe that our argument pertains specifically to sequential stag hunt games, and not to sequential games in general. We do not know to what extent the argument generalizes. For example, we have not been able to ascertain whether the presence of a framing effect in the multi-round relative of the trust game (form) considered by Burnham, McCabe, and Smith (2000) is consistent with a coordination argument. Moreover, and perhaps more importantly, we think that equilibrium arguments, while always problematic in one-shot situations, are even less appropriate in complicated settings such as theirs.

2, except moves are sequential instead of simultaneous. Moreover, in order to maximize statistical power we ask the second mover to report a contingent strategy; one choice in case the first mover plays C and one choice in case the first mover plays D. That is, we adopt the strategy method (Selten, 1967). While we recognize that the strategy method by itself could have a dampening effect on subjects' willingness to reciprocate (Casari and Cason, 2009), there is no immediate reason to expect that the strategy method should also affect the impact of social framing.

In total, 272 subjects participated as decision-makers in the experiment, which was conducted in September 2009. As in Study 2, all were freshmen enrolled in a basic microeconomics course at SSE. Although they come from a later cohort, we thus expect them to have similar characteristics to the SSE students of studies 1 and 2.

Figure 5 displays the proportions of first-movers that choose to play A (i.e., to cooperate) under each of the two social frames in the third study. As expected, the level of cooperation is higher than in the case of simultaneous moves. However, there is no significant social framing effect ($z=0.316$, $p=0.376$). Provided that the expectation about player 2's behavior is at least as optimistic under the Community Game frame as under the Stock Market Game frame, this evidence contradicts the appropriateness hypothesis.

FIGURE 5 ABOUT HERE

Since player 2 can condition the action on player 1's move, there is no role for beliefs when we interpret player 2's behavior. Player 2's action thus provides an even stronger test of the appropriateness hypothesis. Figure 6 displays the results. The first pair of staples denotes, for the Stock Market Game and the Community Game respectively, the fraction of subjects in the role of player 2 that cooperate if player 1 cooperates. The second pair of staples gives the corresponding cooperation rates for the case in which player 1 defects.

FIGURE 6 ABOUT HERE

While there is more cooperation in the Community Game than in the Stock Market game, the difference is minor and not statistically significant (conditional on player 1 cooperating, $z=0.057$, $p=0.477$; conditional on player 1 defecting, $z=0.511$, $p=0.305$).

FIGURE 7 ABOUT HERE

The lack of significant differences is further emphasized in Figure 7, where we break down the observations further and consider all the four strategies that player 2 may use; CC denotes unconditional cooperation (i.e., always playing C), CD denotes conditional cooperation (play of C in response to C and D in response to D), DD denotes unconditional defection, and DC denotes defection in response to cooperation and cooperation in response to defection.²² As the figure shows, the differences across the two treatments are minor. The null hypothesis that the two distributions are identical cannot be rejected (Pearson chi-square = 0.789, $p=0.852$). In the case of cooperation by player 1 there is no difference at all. In the case of defection by player 1, cooperation by player 2 is somewhat more frequent in the Community treatment, but the effect is statistically insignificant ($p=0.192$).

In our view, the last experiment rejects the hypothesis that framing effects come exclusively through the preference channel, at least as we have modeled it.

However, we acknowledge the possibility that the insignificant aggregate effect at the second stage is due to interaction between several types of behavior, of which “appropriateness” is one.²³ For example, suppose that a fraction of the subjects are conditionally altruistic, with their degree of altruism depending on their beliefs about the opponent’s altruism. Suppose that these subjects believe that the logic of appropriateness may be applying to their opponent. Then, in the role of Player 2, they are more prone to reward cooperation under the Stock Market frame (when only the most altruistic opponent types will cooperate) than under the Community frame. Thus, the aggregate insensitivity of Player 2 behavior could be generated by a mix of such conditionally cooperative types and altruistic types responding to the frame. However, the problem with this argument is what it implies for Player 1 equilibrium behavior: In equilibrium, Player 1 behavior must differ across frames, with more cooperation under the Community frame, otherwise conditional altruists will not cooperate more under the Stock Market frame. The lack of a significant difference in Player 1 behavior thus speaks against this interpretation of Player 2 behavior.

²² The strategy DC may seem unintuitive, but actually makes good sense for an unconditional altruist. If the opponent cooperates, own cooperation means giving up 30 in order to give 45. If the opponent defects, own cooperation means giving up 15 in order to give 60.

²³ We are grateful to Jean Tirole and Dirk Engelmann for helpful discussions concerning the ensuing argument.

6. CONCLUSION

Our first experiment demonstrates that situational labels significantly affect behavior in social dilemma situations even under the kind of experimental conditions conventionally imposed by behavioral economists. In this respect, we confirm previous findings that people cooperate more when the name of the game emphasizes the community rather than the individual's interest.

However, the presence of social framing effects does not prove that preferences are unstable or that they depend directly on situational labels. Taken together, our three experiments – involving more than a thousand subjects altogether – instead suggest that social frames primarily serve as coordination devices. These findings are good news for economists. With the preference part of economic models having swelled considerably in recent decades, utility functions arguably carry too heavy an analytical burden already.

Among the five explicit models that we consider, a simple version of the social preferences proposed by Charness and Rabin (2002) is the only model not to be rejected by our experimental data. That is, this formulation seems to be the best description of the game that, to our subjects, corresponds to the Prisoners' dilemma game form. Given such a frame-free description of the game, we conjecture that existing models of belief-based framing effects, notably the level-k model of Bacharach and Stahl (2000), can be straightforwardly adapted to articulate more precisely how framing affect coordination. However, such a complete non-equilibrium analysis is beyond the scope of this paper.

An apparently plausible objection to our conclusion is that we have only rejected a specific formalization of the appropriateness hypothesis. For example, the argument goes, maybe our experimental subjects feel that the social frame only matters to their preferences in the simultaneous version of the Prisoners' dilemma and not in the sequential version? But this objection is flawed. The appropriateness hypothesis has been thriving on rejections of sharply formulated classical decision theory. But to provide a viable alternative, the appropriateness hypothesis must itself be sharply enough formulated to be testable. That is, it must be articulated in a way so as to yield straightforward implications that can be experimentally rejected. While such an improved model may emerge in the future, we conclude that the logic of appropriateness does not explain social framing effects as of now.

It is clearly desirable to investigate the robustness of our findings. For example, one might vary the social frame in other simple game forms, and study other subject pools and contexts. In ongoing work, we are therefore measuring social framing effects in the Dictator game (form). If present, such effects would resuscitate the appropriateness hypothesis, since our preferred model leaves no room for coordination in a game with only a single active player.²⁴ However, in support of the present analysis, the data indicate no social framing effects in the Dictator game.

REFERENCES

Andreoni, J. (1995). "Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments," *Quarterly Journal of Economics* 110(1), 1-21.

Andreoni, J. and Bernheim, B.D. (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects," *Econometrica* 77, 1607-1636.

Bacharach, M. (1993). "Variable universe games," in K. Binmore, A. Kirmanand, and P. Tani (eds.), *Frontiers of Game Theory*, 255-276, Cambridge MA: MIT Press.

Bacharach, M. (1999). "Interactive team reasoning: A contribution to the theory of cooperation," *Research in Economics* 53, 117-147.

Bacharach, M. and Bernasconi, M. (1997). "The variable frame theory of focal points: An experimental study," *Games and Economic Behavior* 19, 1-45.

Bacharach, M. and Stahl, D.O. (2000). "Variable frame level-n theory," *Games and Economic Behavior* 33, 220-246.

Bardsley, N., Mehta, J., Starmer, C., and Sugden, R. (2010). "Explaining focal points: Cognitive hierarchy theory versus team reasoning," *Economic Journal* 120, 40-79.

²⁴ Signaling models of Dictator game behavior, such as Andreoni and Bernheim (2009), do have multiple Perfect Bayesian Nash equilibria.

Batson, C.D. and Moran, T. (1999). "Empathy-induced altruism in a prisoner's dilemma," *European Journal of Social Psychology* 29, 909-924.

Becker, G. (1974). "A theory of social interactions," *Journal of Political Economy* 82, 1063-1093.

Bénabou, R. and Tirole, J. (2006). "Incentives and prosocial behavior," *American Economic Review*, 96(5): 1652-1678.

Bernheim, B.D. (1994). "A theory of conformity." *Journal of Political Economy*, 102(5): 841-877.

Bohnet, I. and Cooter, R.D. (undated). "Expressive law: Framing or equilibrium selection," manuscript, Harvard University.

Bolton, G. and Ockenfels, A. (2000). "ERC: A theory of equity, reciprocity, and competition," *American Economic Review* 90, 166-193.

Brandts, J. and Schwiore, C. (2009). "Frames and economic behavior: An experimental study," manuscript, Universitat Autònoma de Barcelona.

Brennan, G. and Pettit, P. (2004). *The Economy of Esteem*. Oxford: Oxford University Press.

Brewer, M.B. and Kramer, R.M. (1986). "Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing," *Journal of Personality and Social Psychology*, 50(3), 543-549.

Broberg, T., Ellingsen, T., and Johannesson, M. (2007). "Is generosity involuntary?" *Economics Letters* 94, 32-37.

Burnham, T., McCabe, K.A., and Smith, V.L. (2000). "Friend-or-foe intentionality priming in an extensive form trust game," *Journal of Economic Behavior and Organization* 43, 57-73.

Cachon, G.P. and Camerer, C.F. (1996). "Loss-avoidance and forward induction in experimental coordination games," *Quarterly Journal of Economics* 111, 165–194.

Camerer, C.F. (2003) *Behavioral Game Theory: Experiments in Strategic Interaction*. New Jersey: Princeton University Press.

Casari, M. and Cason, T.N. (2009). "The strategy method lowers measured trustworthy behavior," *Economics Letters* 103, 157-159.

Charness, G. and Rabin, M. (2002). "Understanding social preferences with simple tests," *Quarterly Journal of Economics* 117, 817-869.

Chen, Y. and Li, S.X. (2009). "Group identity and social preferences," *American Economic Review* 99, 431-457.

Clark, K. and Sefton, M. (2001). "The sequential Prisoners' dilemma: Evidence on reciprocity," *Economic Journal* 111, 51-68.

Cookson, R. (2000). "Framing effects in public goods experiments," *Experimental Economics* 3, 55-79.

Crawford, V.P., Gneezy, U., and Rottenstreich, Y. (2008). "The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures," *American Economic Review* 98, 1443-1458.

Croson, R.T.A. (1999). "The disjunction effect and reason-based choice in games," *Organizational Behavior and Human Decision Processes* 80, 118-133.

Costa-Gomes, M.A. and Weizsäcker, G. (2008). "Stated beliefs and play in normal-form games," *Review of Economic Studies* 75, 729-762.

Dana, J., Cain, D.M., and Dawes, R.M. (2006). "What you don't know won't hurt me: Costly but quiet exit in dictator games," *Organizational Behavior and Human Decision Processes*, 100, 193-201.

Dawes, R.M., McTavish, J., and 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.

Deutsch, M. (1958). "Trust and suspicion," *Journal of Conflict Resolution* 2, 265-279.

Deutsch, M. (1960). "The effect of motivational disposition on trust and suspicion," *Human Relations* 13, 123-139.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2010). "The framing of games and the psychology of play," University of Nottingham, Cedex Discussion Paper No. 2010-16.

Eiser, J.R. and Bhavnani, K. (1974). "The effect of situational meaning on the behaviour of subjects in the Prisoner's Dilemma Game," *European Journal of Social Psychology* 4(1), 93-97.

Ellingsen, T. and Johannesson, M. (2007). "Paying respect," *Journal of Economic Perspectives* 21(4), 135-149.

Ellingsen, T. and Johannesson, M. (2008). "Pride and prejudice: The human side of incentive theory," *American Economic Review* 98, 990-1008.

Fehr, E. and Schmidt, K.M. (1999). "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics* 114, 817-868.

Fehr, E. and Schmidt, K. (2006). "The economics of fairness, reciprocity and altruism – experimental evidence and new theories," in S.-C. Kolm and J.M. Ythier (eds.): *Handbook of the Economics of Giving, Altruism and Reciprocity*, volume 1. Amsterdam: Elsevier.

Ferraro, F., Pfeffer, J., and Sutton, R. I. (2005). "Economics language and assumptions: How theories can become self-fulfilling," *Academy of Management Review* 30, 8-24.

Frey, B.S. and Osterloh, M. (2005). "Yes managers should be paid like bureaucrats," *Journal of Management Inquiry* 14, 96-111.

Gibbard, A. (1973). "Manipulation of voting schemes: A general result," *Econometrica* 41 (4), 587-601.

Glazer, A., and Konrad, K.A. (1996). "A signaling explanation of charity." *American Economic Review* 86(4), 1019-1028.

Ireland, N. (1994). "On limiting the market for status signals." *Journal of Public Economics* 53(1): 91-110.

Kay, A.C. and Ross, L. (2003) "The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner's dilemma," *Journal of Experimental Social Psychology* 39, 634-643.

Koneberg, C., Yaish, M., and Stocké, V. (2010). "Norms and rationality in electoral participation and in the rescue of Jews in WWII: An application of the model of frame selection," *Rationality and Society* 22 (1), 3-36.

Kreps, D.M., Milgrom, P., Roberts, J., and Wilson, R. (1982). "Rational cooperation in the finitely repeated Prisoners' dilemma," *Journal of Economic Theory* 27 (2), 245-252.

Krupka, E. and Weber, R.A. (2010). "Identifying social norms using coordination games: Why does dictator game sharing vary?" Manuscript, Carnegie Mellon University.

Larrick, R.P and Blount, S. (1997). "The claiming effect: Why players are more generous in social dilemmas than in Ultimatum games," *Journal of Personality and Social Psychology* 72, 810-825.

Lazear, E.P., Malmendier, U., and Weber, R.A. (2010). "Sorting, prices, and social preferences," Manuscript, University of California, Berkeley.

Levin, I.P., Schneider, S.L., and Gaeth, G.J. (1998). "All frames are not created equal: A typology and critical analysis of framing effects," *Organizational Behavior and Human Decision Processes* 76(2), 149-188.

Levine, D.K. (1998). "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics* 1, 593-622.

Levitt, S. and List, J. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21(2), 151-174.

Liberman, V., Samuels, S.M., and Ross, L. (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* 30(9), 1175-1185.

López-Pérez, R. (2008). "Aversion to norm-breaking: A model," *Games and Economic Behavior* 64, 237-267.

March, J. (1994). *A Primer on Decision-Making: How Decisions Happen*. New York: Free Press.

McCabe, K.A., Rigdon, M.L., and Smith, V.L. (2003). "Positive intentions and intentions in trust games," *Journal of Economic Behavior and Organization* 52, 267-275.

Mehta, J., Starmer, C., and Sugden, R. (1994). "The nature of salience: An experimental investigation," *American Economic Review* 84, 658-673.

Montgomery, J.D. (1998). "Toward a role-theoretic conception of embeddedness," *American Journal of Sociology* 104(1), 92-125.

Pfeffer, J. (2007). "Human resources from an organizational behavior perspective: Some paradoxes explained," *Journal of Economic Perspectives* 21(4), 115-134.

Pillutla, M.M. and Chen, X. (1999). "Social norms and cooperation in social dilemmas: The effects of context and feedback," *Organizational Behavior and Human Decision Processes* 78(2), 81-103.

Rabin, M. (1993). "Incorporating fairness into game theory and economics," *American Economic Review* 83, 1281-1302.

Rabin, M. (1998). "Psychology and economics," *Journal of Economic Literature* 36 (1), 11-46.

Radner, R. (1986). "Can bounded rationality resolve the Prisoners' dilemma?" Chapter 20 of A. Mas-Colell and W. Hildenbrand, eds., *Contributions to Mathematical Economics*, Amsterdam: North-Holland.

Rege, M. and Telle, K. (2004). "The impact of social approval and framing on cooperation in public good situations," *Journal of Public Economics* 88, 1625-1644.

Rydval, O. and Ortmann, A., (2005). "Loss avoidance as selection principle: Evidence from simple stag-hunt games," *Economics Letters* 88, 101-107.

Schelling, T.C. (1960). *The Strategy of Conflict*, Cambridge MA: Harvard University Press.

Selten, R. (1967): "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments", *Beiträge zur experimentellen Wirtschaftsforschung*, H. Sauermann (ed.), Vol. I, Tübingen: J.C.B. Mohr (Siebeck), 136-168.

Sen, A.K. (1967). "Isolation, assurance, and the social rate of discount," *Quarterly Journal of Economics* 81, 112-124.

Sobel, J. (2005). "Interdependent preferences and reciprocity," *Journal of Economic Literature* 43, 392-436.

Stahl, D.O. and Wilson, P.W. (1995). "On players' models of other players: Theory and experimental evidence" *Games and Economic Behavior* 10(1), 33-51.

Sugden, R. (1993). "Thinking as a team: Towards an explanation of nonselfish behavior," *Social Philosophy and Policy* 10, 69-89.

Tversky, A. and Kahneman, D. (1981). "The framing of decisions and the psychology of choice," *Science* 211, 453-458.

Tyran, J.-R. and Feld, L. (2006). "Achieving compliance when legal sanctions are non-deterrent," *Scandinavian Journal of Economics* 108(1), 135-56.

van Dijk, E. and Wilke, H. (2000). "Decision-induced focusing in social dilemmas: Give-some, keep-some, take-some, and leave-some dilemmas," *Journal of Personality and Social Psychology* 78(1), 92-104.

Weber, J.M. et al. (2004). "A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness," *Personality and Social Psychology Review* 8(3), 281-307.

Yamagishi, T. and Kiyonari, T. (2000). "The group as the container of generalized reciprocity," *Social Psychology Quarterly* 63, 111-132.

Zhong, C.-B., Loewenstein, J., and Murnighan, J.K. (2007). "Speaking the same language: The cooperative effects of μ in the Prisoners' Dilemma," *Journal of Conflict Resolution* 51(3), 431-456.

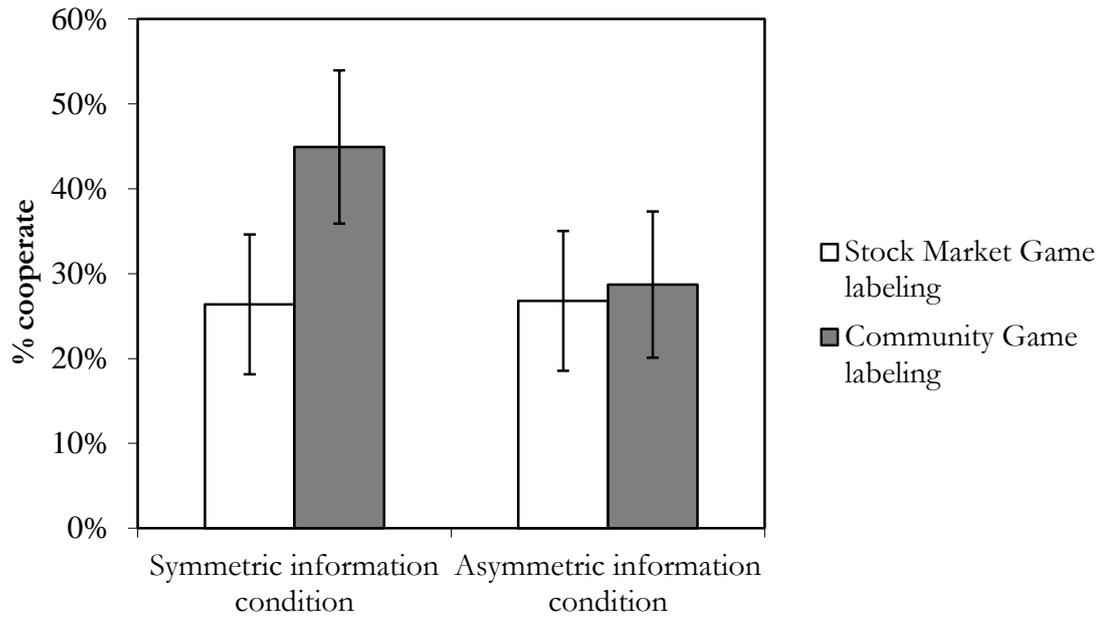


Figure 3. Fraction of cooperative actions in treatments 1-4.

The difference between the first two staples is the social framing effect in the standard symmetric information condition. The difference between the second two staples is the social framing effect in the asymmetric information condition. Error bars indicate 95% confidence intervals.

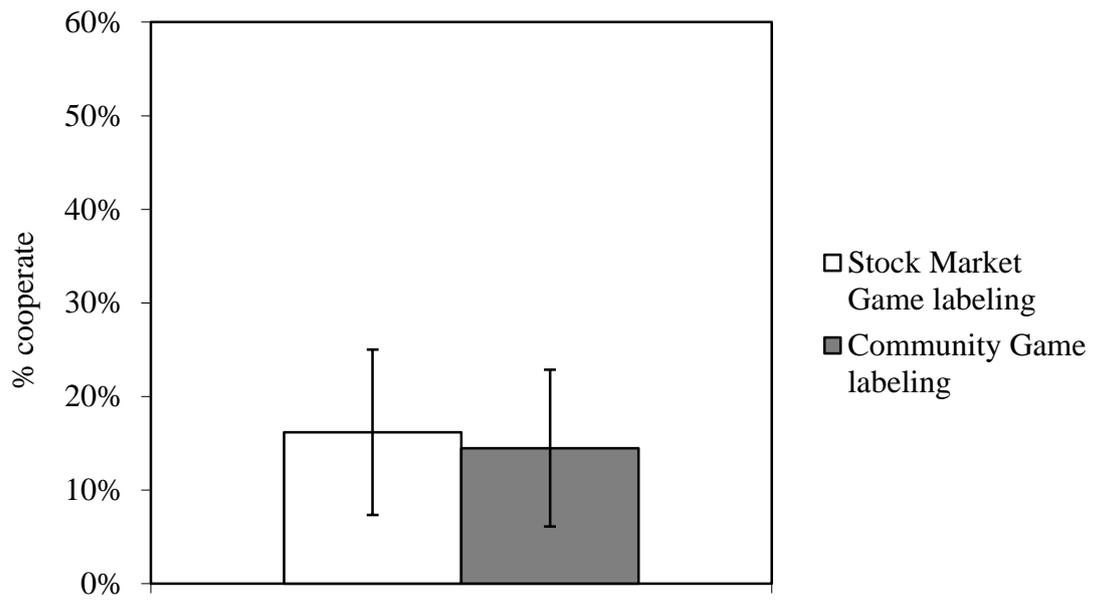


Figure 4. Fraction of cooperative actions in the second experiment, treatments 1' and 2'

The difference between the two staples is the social framing effect in the symmetric information passive opponent condition. Error bars indicate 95% confidence intervals.

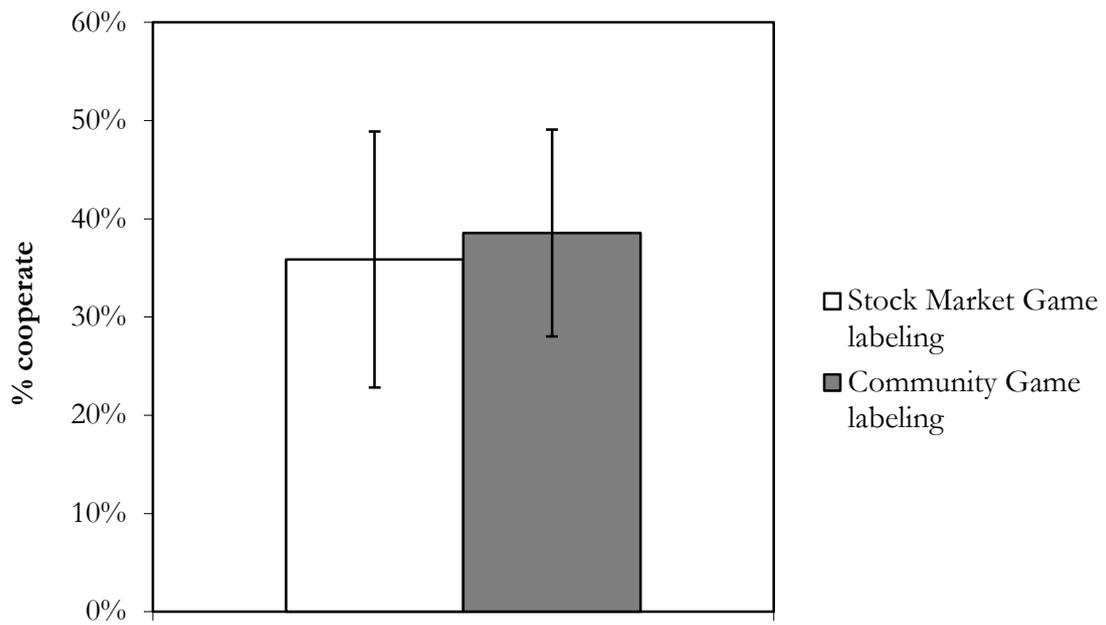


Figure 5. Fraction of first-mover cooperative actions in the third experiment treatments 1'' and 2''

The difference between the two staples is the social framing effect for player 1 in the sequential moves condition. Error bars indicate 95% confidence intervals.

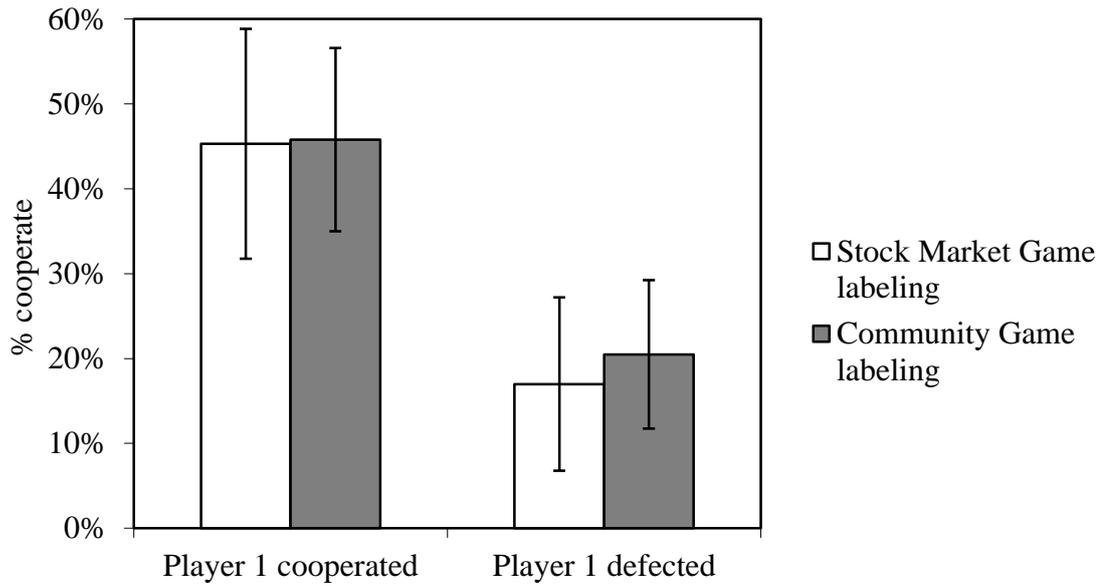


Figure 6. Fraction of second-mover cooperative actions in the third experiment treatments 1'' and 2''

The difference between the first (second) two staples is the social framing effect for player 2, when player 1 cooperated (defected), in the sequential moves condition. Error bars indicate 95% confidence intervals.

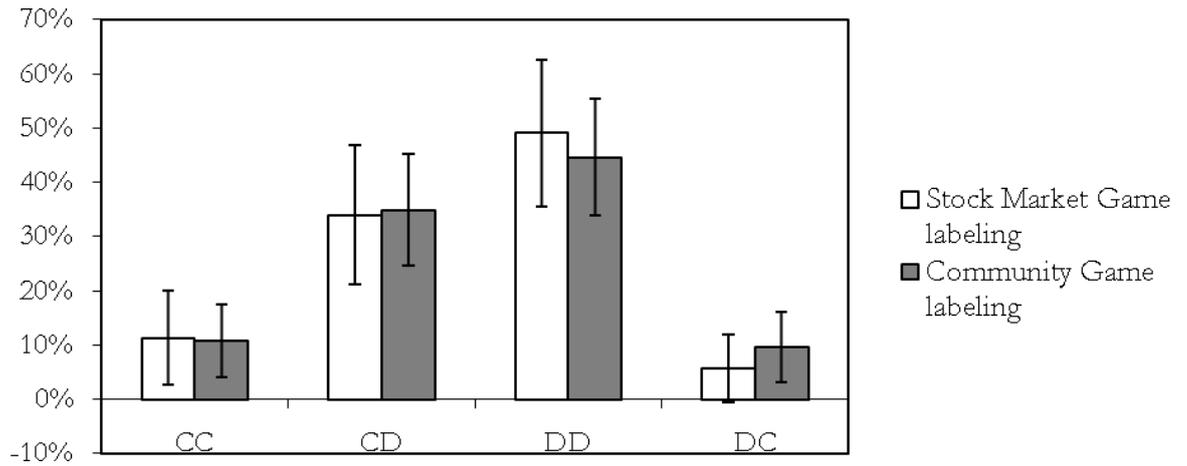


Figure 7. Distribution of second-mover strategies in the third experiment treatments 1'' and 2''

Each staple indicates the fraction of second movers that chose the strategy in the Stock Market frame (first staple in each pair) and the Community frame (second staple) respectively. Error bars indicate 95% confidence intervals.

APPENDIX I: INSTRUCTIONS FOR THE PARTICIPANTS IN TREATMENT 1 AND 2 IN STUDY 1²⁵

Stock Market Game [Community Game]²⁶, instructions

Hi and welcome. You are going to take part in the **Stock Market Game [Community Game]**. For your participation you will get compensation. This compensation is dependent on the choices you make.

Please read the instructions carefully. If you have any questions, please raise your hand and the experimenter will come and help you. Do not ask questions without raising your hand first. It is also important that you do not speak to the other participants while the experiment is taking place.

In the Stock Market Game [Community Game] you are paired up with a person in another room. You will not get any information about who that person is, neither before nor after the experiment. The other person will not get information about your identity either.

All people in this room and all people in the other room get the same instructions and compensations for taking part in the experiment.

The Stock Market Game [Community Game] looks like this:

Stock Market Game [Community Game]

		You	
		A	B
The other person	A	50 50	5 80
	B	80 5	20 20

You and the other person choose simultaneously between A and B. Depending on your respective choices, you end up in one of the four squares in the matrix above. The bold numbers in the upper right corner represents, in Swedish kronor, what you get and the numbers in the lower left corner represents what the other person gets.

Examples:

- If both you and the other person choose A you both get 50 kronor.
- If both you and the other person choose B you both get 20 kronor.
- If you choose A and the other person chooses B you get 5 kronor and the other person gets 80 kronor.
- If you choose B and the other person chooses A you get 80 kronor and the other person gets 5 kronor.

²⁵ Appendices B, C and D are not intended for publication.

²⁶ The name ‘Stock Market Game’ was used in treatment 1 and the name ‘Community Game’ was used in treatment 2.

Please note that you will not know anything about the decision of the other person when you make your decision.

Write your decision on the form marked “answering form”. Then turn the form upside-down and put it in front of you.

When the Stock Market Game [Community Game] is finished, the experimenter will compile the results and prepare an envelope for each participant. These envelopes will then be distributed. The envelope will contain information about how the other person decided and what the result of the game was. The sum you are allotted will also be in the envelope.

Thank you for your participation!

APPENDIX II: INSTRUCTIONS FOR THE PARTICIPANTS IN TREATMENT 3 AND 4 IN STUDY 1

Stock Market Game [Community Game]²⁷, instructions

Hi and welcome. You are going to take part in the **Stock Market Game [Community Game]**. For your participation you will get compensation. This compensation is dependent on the choices you make.

Please read the instructions carefully. If you have any questions, please raise your hand and the experimenter will come and help you. Do not ask questions without raising your hand first. It is also important that you do not speak to the other participants while the experiment is taking place.

In the Stock Market Game [Community Game] you are paired up with a computer. To the computer a receiver is connected. The receiver is a person sitting in another room. You will not get any information about who the receiver is, neither before nor after the experiment. The receiver will not get information about your identity either.

The receiver makes no decision during the experiment but get the compensation that the computer is allotted. This person knows only that the money that she or he gets is the results of an experiment. The person does not know which game that has been played and thus not how you or the computer acted.

All people in this room get the same instructions and compensations for taking part in the experiment.

²⁷ The name ‘Stock Market Game’ was used in treatment 3 and the name ‘Community Game’ was used in treatment 4.

The Stock Market Game [Community Game] looks like this:

Stock Market Game [Community Game]

		You	
		A	B
The computer / the receiver	A	50 50	5 80
	B	80 5	20 20

You and the computer choose simultaneously between A and B. Depending on your respective choices, you end up in one of the four squares in the matrix above. The bold numbers in the upper right corner represents, in Swedish kronor, what you get and the numbers in the lower left corner represents what the receiver gets.

Examples:

- If both you and the computer choose A you get 50 kronor and the receiver gets 50 kronor.
- If both you and the computer choose B you get 20 kronor and the receiver gets 20 kronor.
- If you choose A and the computer chooses B you get 5 kronor and the receiver gets 80 kronor.
- If you choose B and the computer chooses A you get 80 kronor and the receiver gets 5 kronor.

When the computer chooses between A and B it is done in the following way: we conduct this experiment also with people playing against each other. Depending on how the players act in that game we calculate with which probability the computer must choose A and B respectively to “imitate” the behavior of a human player.

Please note that you will not know anything about the decision of the computer when you make your decision.

Write your decision on the form marked “answering form”. Then turn the form upside-down and put it in front of you.

When the Stock Market Game [Community Game] is finished, the experimenter will compile the results and prepare an envelope for each participant in this room. These envelopes will then be distributed. The envelope will contain information about how the other person decided and what the result of the game was. The sum you are allotted will also be in the envelope. The receivers in the other room will also get an envelope with the money that the computer was allotted, but no further information.

Thank you for your participation!

APPENDIX III: INSTRUCTIONS FOR THE PARTICIPANTS IN STUDY 2

STOCK MARKET GAME [COMMUNITY GAME]²⁸, INSTRUCTIONS

Hi and welcome. You are going to take part in the **stock market game [community game]**. For your participation you will get compensation. This compensation is dependent on the choices you make.

Please read the instructions carefully. If you have any questions, please raise your hand and the experimenter will come and help you. Do not ask questions without raising your hand first. *It is also important that you do not speak to the other participants while the experiment is taking place!*

In the stock market game [community game] you are paired up with a computer. To the computer a receiver is connected. The receiver is a person sitting in another room. You will not get any information about who the receiver is, neither before nor after the experiment. The receiver will not get information about your identity either.

The receiver makes no decision during the experiment but get the compensation that the computer is allotted. The receiver has received a copy of these instructions and has been told orally that he/she will receive the compensation that the computer is allotted.

The stock market game [community game] looks like this:

Stock market game [Community game]

		You	
		A	B
The computer / the receiver	A	50 50	5 80
	B	80 5	20 20

You and the computer choose simultaneously between A and B. Depending on your respective choices, you end up in one of the four squares in the matrix above. The bold numbers in the upper right corner represents, in Swedish kronor, what you get and the numbers in the lower left corner represents what the receiver gets.

Examples:

- If both you and the computer choose A you get 50 kronor and the receiver gets 50 kronor.
- If both you and the computer choose B you get 20 kronor and the receiver gets 20 kronor.
- If you choose A and the computer chooses B you get 5 kronor and the receiver gets 80 kronor.

²⁸ The name 'stock market game' was used in treatment 1' and the name 'community game' was used in treatment 2'.

- If you choose B and the computer chooses A you get 80 kronor and the receiver gets 5 kronor.

When the computer chooses between A and B it is done in the following way: the choice is done randomly with the same probability that the computer will choose A as B (i.e. like flipping a coin between option A and B).

Please note that you will not know anything about the decision of the computer when you make your decision.

Write your decision on the form marked “answering form”. Then turn the form upside-down and put it in front of you.

When the stock market game [community game] is finished, the experimenter will compile the results and prepare an envelope for each participant. These envelopes will then be distributed. The envelope will contain information about how the computer decided and what the result of the game was. The sum you are allotted will also be in the envelope. The receivers in the other room will also get an envelope with the money that the computer was allotted, and information about your choice and the choice of the computer.

Thank you for your participation!

APPENDIX IV: INSTRUCTIONS FOR THE PARTICIPANTS IN STUDY 3

STOCK MARKET GAME [COMMUNITY GAME]²⁹, INSTRUCTIONS

Hi and welcome. You are going to take part in the **stock market game [community game]**. For your participation you will get compensation. This compensation is dependent on the choices you make.

Please read the instructions carefully. If you have any questions, please raise your hand and the experimenter will come and help you. Do not ask questions without raising your hand first. *It is also important that you do not speak to the other participants while the experiment is taking place!*

In the stock market game [community game] you are paired up with a person in another room. The persons in one of the rooms are called player 1 and the persons in the other room are called player 2. You and everyone else in your room are **player 1 (2)**. You will not get any information about who the person in the other room is and he/she will not get any information about your identity, neither before nor after the experiment. All the persons in both rooms have received these instructions.

The stock market game [community game] looks like this:

Stock market game [Community game]

		Player 1	
		A	B
<i>Player 2</i>	A	50 50	5 80
	B	80 5	20 20

Player 1 first chose between A and B without knowing which choice player 2 will make. Thereafter player 2 chose between A and B, given that she knows which choice player 1 has made. The bold numbers in the upper right corner represents, in Swedish kronor, what player 1 get and the numbers in italics in the lower left corner represents what player 2 gets.

Examples:

- If both player 1 and player 2 choose A you both get 50 kronor.
- If both player 1 and player 2 choose B you both get 20 kronor.
- If player 1 choose A and player 2 choose B, player 1 get 5 kronor and player 2 gets 80 kronor.
- If player 1 choose B and player 2 choose A, player 1 get 80 kronor and player 2 gets 5 kronor.

²⁹ The name ‘stock market game’ was used in treatment 1’’ and the name ‘community game’ was used in treatment 2’’.

Player 1 write down their choice on the form marked “answering form (player 1)” and player 2 write down their choice on the form marked “answering form (player 2)”. Player 2 writes down their choice both for the case when player 1 chose A and for the case when player 1 chose B (the choice of player 2 for the choice actually made by player 1 will then be used to determine the payments). When you have made your choice, turn the form upside-down and put it in front of you.

The experimenters then collect the forms in both rooms and prepare an envelope for each participant that contains the payment of each participant. These envelopes will then be distributed and the experiment is then over.

Thank you for your participation!