# Social Framing Effects

## Why Do Situational Labels Affect Cooperation?

Tore Ellingsen
Stockholm School of Economics

Magnus Johannesson
Stockholm School of Economics

Johanna Mollerstrom
Harvard University

Sara Munkhammar
Ramböll Management Consulting

This version: 23 June 2010

**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, David Cesarini, Vincent Crawford, Ernst Fehr, Michael Kosfeld, John List, Craig McKenzie, Erik Mohlin, Martin Nowak, Charlie Plott, Björn Wallace, Robert Östling, and many seminar participants for helpful 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

People's behavior responds significantly to situational labels. In a seminal experiment, 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).[1]

These findings have 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 seems 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 employee effort (Frey and Osterloh, 2005; Pfeffer, 2007), legal compliance (Tyran and Feld, 2006; Bohnet and Cooter, undated), and other prosocial 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 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

[1] Other studies that investigate the impact of labels in social dilemmas include, *inter alia*, Brewer and Kramer (1986), Andreoni (1995), 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.

based on the "logic of appropriateness," as articulated by March (1994, Chapter 2), we refer to it as the *appropriateness hypothesis.*

A separate but closely related hypothesis is that people respond to labels because the label affects how others interpret their behavior, which in turn determines their social esteem. This is the *social image hypothesis.*[2]

A third class of theories of social framing 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.[3] Since a Stag hunt game has two pure strategy Nash equilibria, the label can be used as an equilibrium selection device, as noted by Rabin (1998) and Fehr and Schmidt (2006). At the outset, we lump these 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, so as to identify their main differences. 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 under both move sequences. 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.

---

[2] 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.

[3] 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).

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.[4] 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.

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. Whereas Liberman, Samuels, and Ross (2004) suggest that the social frame could be affecting behavior through beliefs, they 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.

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

[4] 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.

more in-group favoritism in Prisoners' dilemmas with simultaneous play than in games with sequential play.[5] Although Yamagishi and Kiyonari do not explicitly invoke the game theoretic argument, it is clear that their idea is essentially the same as 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 most convincing evidence comes from the absence of a framing effect among second movers in the sequential game. For second movers expectations play no role, so any effect must be caused by preferences alone.

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 labeling effects exist and that they can be removed by suitable manipulations of the environment. More precisely, the study shows that the framing 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.

[5] 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.

## 2. THEORY

Consider two persons, Rowena and Colin, facing the game form depicted in Figure 1. In each cell, the first number is Rowena's material payoff, and the second number is Colin's material payoff.

| | A | B |
|---|---|---|
| A | *1,1* | *c,d* |
| B | *d,c* | *0,0* |

***Figure 1: Material payoffs - the game form***

Let $d>1$ and $c<0$. Moreover, let $0<c+d<2$. Thus, the sum of the material payoffs is largest if both Rowena and Colin choose action A. However, if Rowena and Colin are selfish materialists, they will both be playing B, since B maximizes the own material payoff regardless of what the opponent does. That is, the game form in Figure 1 is a Prisoners' dilemma.

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.

### 2.1 Frame-dependent preferences

Suppose that Rowena and Colin embrace a social efficiency norm, that is, a norm saying that people ought to choose the action that maximizes the sum of material payoffs. For simplicity, let own norm compliance yield an extra utility corresponding to a material payoff $\eta$.[6] Then, the actual game corresponding to the game form in Figure 1 is given by the bi-matrix depicted in Figure 1a.

| | A | B |
|---|---|---|
| A | $1+\eta,1+\eta$ | $c+\eta,d$ |
| B | $d,c+\eta$ | $0,0$ |

***Figure 1a: The game with an unconditional efficiency norm***

Clearly, if $\eta>\max\{d-1,-c\}$, each player's dominant action is to play A, whereas action B remains dominant if $\eta<\min\{d-1,-c\}$. Observe that beliefs about the opponent are irrelevant whenever a player has a dominant action. Thus, even if people differ with respect to their pro-

[6] Our formulation implies that the efficiency norm is internalized. See, e.g., Andreoni and Bernheim (2009) or Krupka and Weber (2009) for richer models of norm compliance based on internalization.

pensity to comply with norms, the belief about the opponent's propensity does not enter into the decision problem unless $\eta$ falls between $d$-$l$ and -$c$.

A natural way to think of the appropriateness hypothesis is to let $\eta$, the strength of the efficiency norm, depend on the decision frame. 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 drawback of this explanation is that we do not yet have a theory of how social frames may be entering preferences. Certainly, useful theory should avoid having players' preferences depend in complex ways on the game form.

The social esteem hypothesis says that people are concerned about what others may think about them. For example, Colin may like to think that Rowena thinks that Colin obeys social norms. Formally, Colin's belief about Rowena's belief will then enter Colin's utility function. As a result, the Prisoners' dilemma turns into a (two-sided) signaling problem, in which players may cooperate not only because they desire to follow norms, but also in order to *convey the impression* that they follow norms.[7] If the desire to be seen to follow a norm depends on the frame, this hypothesis works in essentially the same way as the appropriateness hypothesis. 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

Suppose now instead that the norm is to prioritize efficiency if and only if others do so as well, and that Rowena and Colin both have internalized the norm. To formalize this conditional efficiency norm, we assume that $\eta$ is positive only in the (A,A) cell. That is, the game is as depicted in Figure 1b.

| | A | B |
|---|---|---|
| A | $1+\eta,1+\eta$ | $c,d$ |
| B | $d,c$ | $0,0$ |

***Figure 1b: The game with a conditional efficiency norm***

If $\eta>d-1$, the game has two pure strategy equilibria, namely (A,A) and (B,B). The former equilibrium always Pareto-dominates the latter. If, in addition, $c$ is sufficiently small,

[7] For early formal models of social esteem, 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).

(B,B) is risk-dominant. That is, the game is not a Prisoners' dilemma, but a Stag hunt. To the best of our knowledge, this argument was first made by Sen (1967).

Recall that any uncertainty about the opponent's behavior has little impact on own behavior under the unconditional efficiency norm. Under the conditional efficiency norm, on the other hand, beliefs about the opponent's type are crucial. If the opponent is believed to be selfish with a large probability, it takes a large value of $\eta$ for unselfish players to risk playing A. Conversely, for any $\eta$ characterizing the most unselfish types, action A is played by these unselfish types only if the fraction of selfish types is believed to be sufficiently small.

There are many variants of the above argument. Models that admit multiple equilibria in a Prisoners' dilemma game form include the conditional altruism model of Levine (1998), the inequity aversion models of Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), the conditional fairness model of López-Pérez (2008), as well as the intention-based fairness model of Rabin (1993).[8]

When the game has multiple equilibria, it is a short step to realize that the name of the game can be used as a coordination device, or focal point (Schelling, 1960). However, in order for the name of the game 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.[9]

Note however that this theory is as incomplete as the preference-based theory, since there is no deep explanation for how frames affect beliefs.

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

Two of the treatments are intended to investigate whether we can replicate previous findings of social framing effects within an experiment that satisfies current standards in be-

[8] Refer to Appendix A for a thorough discussion of Fehr and Schmidt in this context.

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

havioral economics.[10] 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 SEK[11] (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.

| | 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.

[10] For the original experiment, see Kay and Ross, 2003 and Liberman, Samuels, and Ross, 2004.

[11] 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.

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 choice between A and B based on the frequencies of play in treatment 1 (2). 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.

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.

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

### 3.2 Findings

The findings are displayed in Figure 3.[12] They reveal a social framing effect in the symmetric information condition. The fraction of subjects choosing the cooperative choice increases from 26.4% with the Stock Market Game label to 44.9% with the Community Game label. This difference is statistically significant (one-sided z-test; z=2.972, p=0.0015),[13] rejecting the null hypothesis of zero framing effect. Under asymmetric information, on the other hand, the framing effect is highly insignificant with 26.8% choosing the cooperative choice with the Stock Market game label and 28.7% with the Community Game label (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

[12] 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.

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

same under asymmetric information as under symmetric information, and favoring the alternative hypothesis that the framing effect is larger under symmetric information.[14]

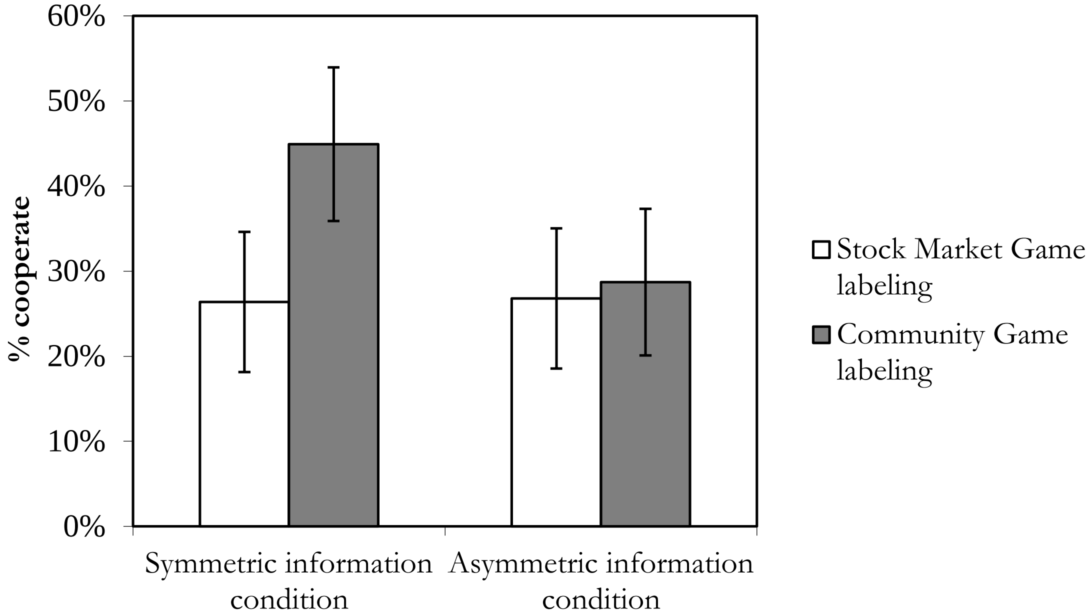


***Figure 3. Fraction of cooperative actions in treatments 1-4.***
*The difference between the first two staples is the social framing effect in the standard symmetric information condition. The difference between the second two staples is the social framing effect in the asymmetric information condition. Error bars indicate 95% confidence intervals.*

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, so we seek reasons why some people may want to cooperate in treatment 2, while not wanting to cooperate in treatment 4.

Two potential reasons for the absence of a framing effect in treatments 3 and 4 spring to mind. First, people may want to cooperate only with others who themselves actively have chosen to cooperate. Then, they will not cooperate when the opponent is unable to choose her own action. Second, people may cooperate primarily to impress the opponent, in which case

[14] 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).

they do not cooperate when the opponent is uninformed about their action. Other reasons can be imagined too. For example, the mere mentioning of a computer can reduce people's urge to behave cooperatively or even affect their subjective beliefs about the matched opponent's action.

Looking at the results in light of the three hypotheses that we discussed in the Introduction tells us the following. First, it rejects a strict version of the appropriateness hypothesis, since the name of the game ceases to matter when subjects meet uninformed and passive opponents.

Second, with respect to the coordination hypothesis, the study rejects all models that are based on the twin assumptions that subjects hold rational "frequentist" beliefs and that they have preferences defined exclusively on material outcomes. In particular, a subject with Fehr-Schmidt inequality aversion and rational frequentist beliefs ought to behave in exactly the same way in treatments 2 and 4 – because by construction the frequency distribution of the opponent's play is the same in the two cases. However, the results do not reject less strict versions of the coordination hypothesis, where agents either do not hold rational beliefs or where also non-material outcomes enter the utility function.[15]

Third, as discussed above, it may be that agents choose to cooperate more in the Community frame than in the Stock Market frame because they care about their social image. The results from our first experiment do not reject that hypothesis.

## 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. That is, even in an anonymous context they want to be "seen" to do what the situation asks of them.[16] The findings of the first study are consistent

[15] In principle, we could have – and possibly should have – investigated which of the two hypotheses are rejected by eliciting beliefs. 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). Recall that Liberman, Samuels and Ross (2004) elicit beliefs both by participants and non-participants in their social framing experiment. While the findings are interesting, they do not speak decisively on the choice between our main hypotheses.

[16] 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).

with this hypothesis, but as noted above it does not allow us to distinguish this social esteem explanation from many others. The second study is designed to provide a sharper test.

In the second study, we modify treatments 3 and 4 of the first study in one crucial respect: We let 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 now ought to be a similar gap between treatments 1' and 2' as between treatments 1 and 2. In order to maximize statistical power for the test between treatments 1' and 2', we refrain from re-running the first study with the new subjects.

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 14, but the baseline level of cooperation (action A) is lower. Therefore, any difference in framing effects is unlikely to be caused by subject pool effects.

Figure 4 displays the findings.

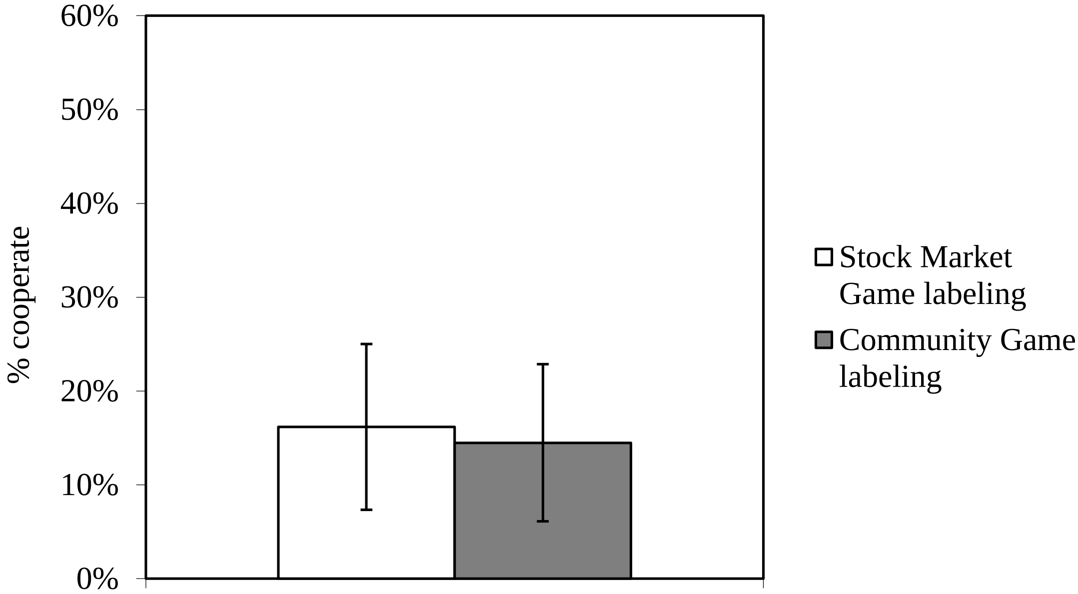


***Figure 4. Fraction of cooperative actions in the second experiment, treatments 1' and 2'***
*The difference between the two staples is the social framing effect in the symmetric information passive opponent condition. Error bars indicate 95% confidence intervals.*

The difference in behavior across treatments is highly insignificant (z=-0.271, p=0.393), and the point estimate even has the wrong sign. Therefore, we reject the hypothesis that the social framing effect in Study 1 was caused by social esteem considerations.

Study 2 casts severe doubt on the hypothesis that social esteem concerns, our third hypothesis from the Introduction, are responsible for the social framing effects in Study 1. To what extent does the evidence also speak to the other hypotheses? As we show in Appendix A, the joint findings of Studies 1 and 2 are consistent with a conditional fairness model, but inconsistent with a conditional altruism model. However, Study 2 offers no new evidence that helps us distinguish more broadly between the appropriateness hypothesis and the coordination hypothesis.

## 5. The Third Study: Preferences or Coordination?

After running the first two studies it occurred to us that there is a straightforward way to distinguish between the appropriateness hypothesis and the coordination hypothesis, namely by letting the moves be sequential instead of simultaneous.[17] If a Stag hunt game is played sequentially, the second player can always assure herself of playing a best reply, and hence the efficient equilibrium is the unique subgame perfect outcome. Thus, if the coordination hypothesis is correct, there should be no social framing effect. On the other hand, if the ap-

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

propriateness 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 A by playing A 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 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 A and one choice in case the first mover plays B. That is, we adopt the strategy method (Selten, 1967).

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.

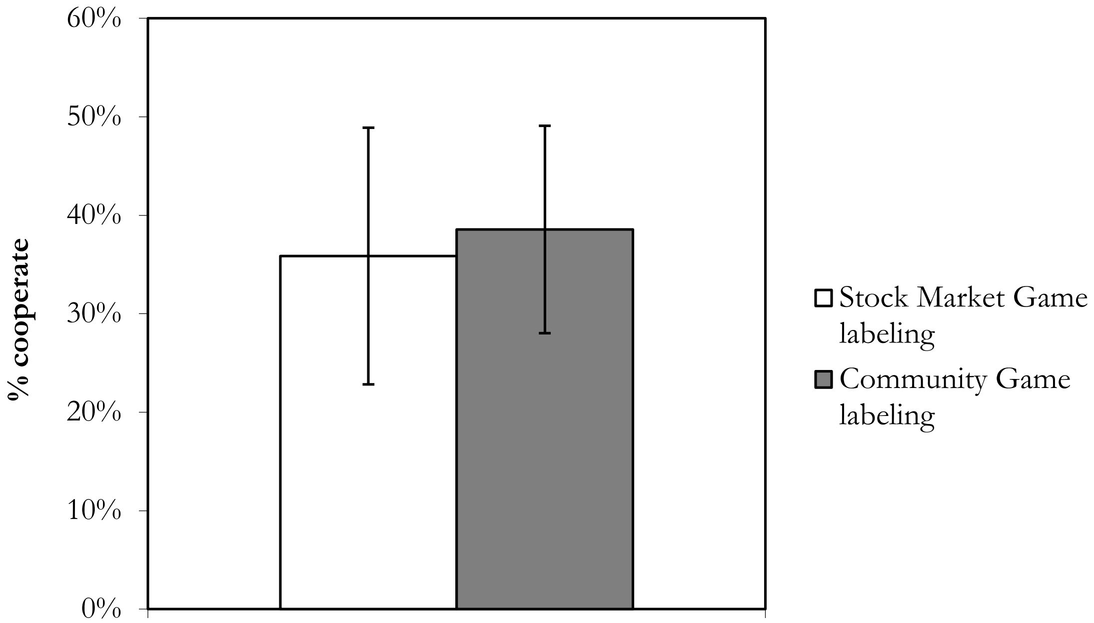


***Figure 5. Fraction of first-mover cooperative actions in the third experiment treatments 1'' and 2''***
*The difference between the two staples is the social framing effect for player 1 in the sequential moves condition. Error bars indicate 95% confidence intervals.*

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.

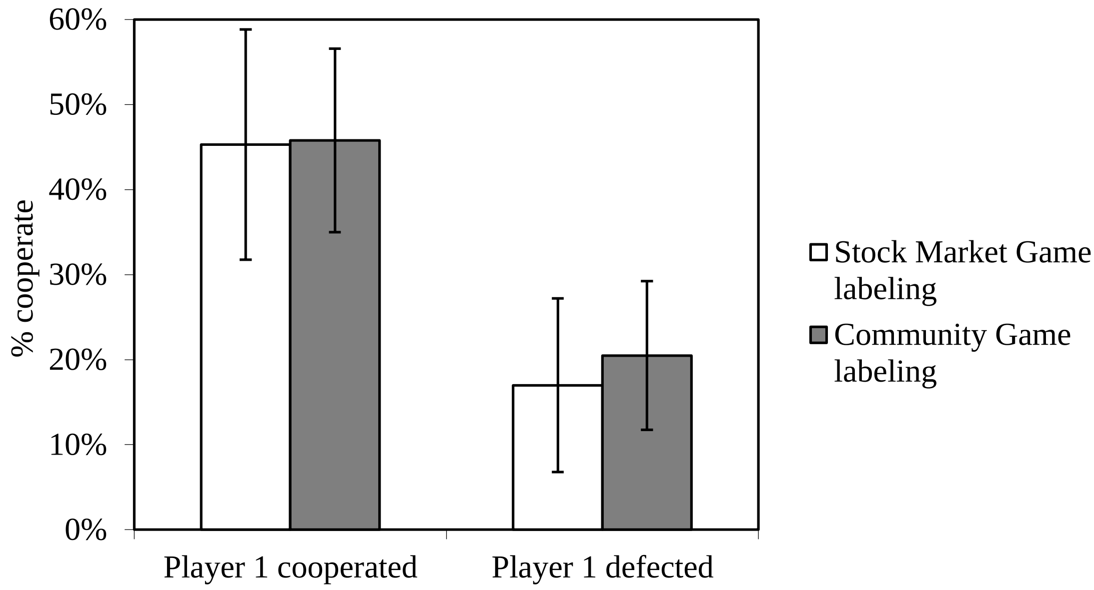


***Figure 6. Fraction of second-mover cooperative actions in the third experiment treatments 1'' and 2''***
*The difference between the first (second) two staples is the social framing effect for player 2, when player 1 cooperated (defected), in the sequential moves condition. Error bars indicate 95% confidence intervals.*

While there is more cooperation in the Community Game than in the Stock Market game, the difference is minor and far from statistically significant (conditional on player 1 cooperating, z=0.057, p=0.477; conditional on player 1 defecting, z=0.511, p=0.305).

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 A), CD denotes conditional cooperation (play of A in response to A and B in response to B), DD denotes unconditional defection, and DC denotes defection in response to cooperation and cooperation in response to defection.[18] As the figure shows, the differences across the two treatments are minor. The null hypothesis that the two distributions are identical is accepted (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 far from being statistically significant (p=0.192).

[18] 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.

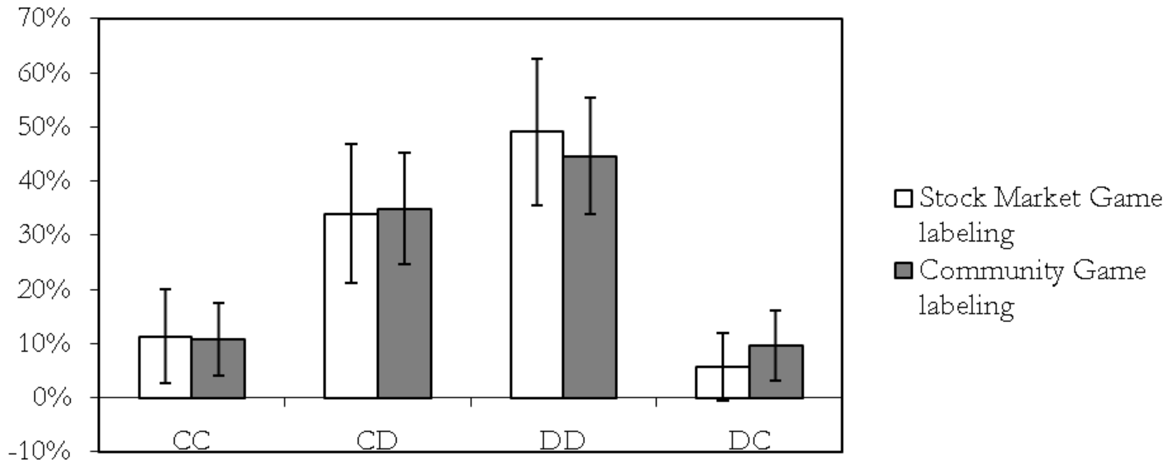


***Figure 7. Distribution of second-mover strategies in the third experiment treatments 1'' and 2''***
*Each staple indicates the fraction of second movers that chose the strategy in the Stock Market frame (first staple in each pair) and the Community frame (second staple) respectively. Error bars indicate 95% confidence intervals.*

## 6. CONCLUSION

We have shown that situational labels significantly affect behavior in social dilemma situations even under the kind of experimental conditions conventionally imposed by behavioral economists. However, the presence of social framing effects does not prove that preferences are unstable or that they depend directly on situational labels. Instead, our analysis suggests that situational labels primarily serve as coordination devices.

In future research the robustness of our findings should be investigated by varying the social frame in other simple game forms. If it can be generally confirmed that social framing affect behavior primarily through beliefs, this will be a useful further step towards providing a theory of framing. However, a complete theory of social framing effects requires an explicit model of belief formation in groups. We are hopeful that our evidence will encourage the development of such models.[19]

[19] One idea is to adapt the level-k model of beliefs developed by Nagel (1995) and Stahl and Wilson (1995). Crawford (2003) and Ellingsen and Östling (2010) have already adapted the level-k model to study the role of pre-play communication.

## 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.

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.

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.

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

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 reciprocation," *Economic Journal* 111, 51-68.

Crawford, V.P. (2003). "Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions," *American Economic Review*, 93(1), 133-149.

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.

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.

Ellingsen, T. and Östling, R. (2010). "When does communication improve coordination?" *American Economic Review* 100 (5), forthcoming.

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.

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

Nagel, R. (1995). “Unraveling in guessing games: An experimental study,” *American Economic Review* 85(5), 1313-1326.

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.

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.

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 labeling in the Prisoners’ Dilemma,” *Journal of Conflict Resolution* 51(3), 431-456.

## APPENDIX A: Fairness preferences

In this section we first show how the Prisoners' dilemma game form may turn into a Stag hunt game if players are inequity (inequality) averse. If beliefs remain constant, such fairness preferences cannot explain why the framing effect is stronger under the active opponent condition than under the passive opponent conditions. However, we go on to show that there exist conditional fairness preferences that can account for the result even under constant beliefs.

Suppose for simplicity that all players have utility functions

$$U_i = m_i - \beta \max \{0, m_i - m_j\}, \qquad \text{(A1)}$$

where $m_i$ denotes player $i$'s material payoff, $m_j$ denotes the opponent's payoff, and $\beta$ is a non-negative parameter. If $\beta > 0$, player $i$ suffers disutility from being better off than the opponent we say that player $i$ is *superiority averse*. Of course, a more common and plausible formulation of inequity aversion arises if we replace $m_i - m_j$ by the absolute value $|m_i - m_j|$, so that players are inferiority averse as well as superiority averse. Indeed, in Fehr and Schmidt's (1999) model, inferiority aversion is stronger than superiority aversion, and the strength of these aversions is heterogeneous across players. Our basic argument can be articulated within that more general model as well, but the analysis gets substantially more complicated.

The utilities arising from our Prisoners' dilemma are then $U_i(A,A) = 50$, $U_i(A,B) = 5$, $U_i(B,A) = 80-75\beta$, and $U_i(B,B) = 20$.

It is straightforward to check that the strategy profile *(B,B)* is a Nash equilibrium. In order for *(A,A)* to also be an equilibrium, we need that $50 > 80-75\beta$, or equivalently that $\beta > 2/5$. Thus, we have illustrated that fairness preferences can transform the Prisoners' dilemma game form into a Stag hunt game. Because there are two pure strategy equilibria, the players may use the labels as a coordination device. That is, the fairness model is rich enough to allow the possibility that labels matter.

However, since the players' utilities depend only on the actions that are taken, the fairness model does *not* allow a player to play any differently in our passive opponent treatments than in the active opponent treatment. To see why, consider a player who plays under the Community game label in the symmetric information condition. Suppose that the player expects the opponent to play strategy *A* with probability $p$ and action *B* with probability $(1-p)$. The player will then unambiguously choose strategy *A* if $50p + 5(1-p) > (80-75\beta)p+20(1-p)$, or equivalently if $\beta > (p+1)/5p$. Conversely the player will choose strategy *B* if $\beta < (p+1)/5p$. Suppose now that the same player were to play the community game under the passive opponent condition. By construction, $p$ is the same. Therefore, the player must take the same action, except possibly in the knife-edge (measure zero) case in which the player expects the opponent to play strategy *A* with probability $p = 1/(5\beta-1)$. In other words, the inequity aversion model can not explain the observed behavior if beliefs are the same in the two treatments. It is straightforward to check that the same argument holds irrespective of the number of player types, that is, irrespective of the support of the $\beta$ distribution. The only crucial assumption is that inequity aversion is unconditional.

Let us now change the model by making inequity aversion conditional, as in López-Pérez (2008). To be precise, let us suppose that there are two player types. One player type is selfish, having inequity aversion $\beta = 0$. The other type is conditionally fair; $\beta = 0$ if the opponent is believed to be selfish, but $\beta > 0$ if the opponent is believed to be conditionally fair. That is, a conditionally fair player does not mind taking advantage of selfish opponents, but dislikes using conditionally fair opponents. Let us assume that the player types have common prior beliefs. Let $x$ denote the (prior) probability of facing a conditionally fair opponent. Note that the selfish type will always find it optimal to play *B*.

In the active opponent condition, it is straightforward to check that *(B,B)* is a Bayesian Nash equilibrium; if a conditionally fair player *i* believes the opponent to play *B*, the best response is to also play *B*, since this maximizes the player *i*'s material payoff and will not generate any superiority costs as long as the opponent plays *B*.

In the active opponent condition, there can also exist equilibria in which conditionally fair players play *A*. The parameter restriction is given by the requirement that a conditionally fair player plays *A* if she believes that every conditionally fair opponent will do so, that is, $50x+5(1-x) > (80-75\beta)x+20(1-x)$, or equivalently $\beta > (x+1)/5x$.

In the passive opponent conditions, a conditionally fair player is less reluctant to play *B*. The reason is that the opponent's play of *A* is no longer a sign that the opponent is fair, since the opponent did not actually choose the action herself. More specifically, the joint probability that the opponent is fair and chooses action *A* is at most $x^2$, not $x$. Thus, there still may exist an equilibrium in which conditionally fair players takes action *A*, but only if $50x^2+50x(1-x)+5x(1-x)+5(1- x)^2 > (80-75\beta) x^2+(80-75\beta)x(1-x)+20x(1-x)+2(1-x)^2$, or equivalently $\beta > (x+1)/5x^2$. For example, suppose $x=2/5$. Then the requirement for multiple equilibria in the active opponent treatment is that $\beta > 7/10$, whereas the requirement in the passive opponent treatments is $\beta > 7/2$. Thus, there is a sizeable set of parameters such that labels can matter under active opponent choice but cannot matter under passive opponent choice.

The fact that a conditional fairness model can fit the evidence does not imply that other models of conditional social preferences can do so. For example, conditional altruism, as defined in Levine (1998), if anything predicts more play of *A* under asymmetric information than under symmetric information. Briefly, the reason is that an altruist who only gets positive utility from helping other altruists is more inclined to help an altruistic opponent by playing *A* if the altruistic opponent is believed to be playing *B* (in which case the "gift" is 80-20=60 and the cost is 20-5=15) than if the altruistic opponent is believed to play *A* (in which case the "gift" is smaller, at 50-5=45 and the cost is larger, at 80-50=30).

## APPENDIX B: INSTRUCTIONS FOR THE PARTICIPANTS IN TREATMENT 1 AND 2 IN STUDY 1[20]

### Stock Market Game [Community Game][21], 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**<br>50 | **80**<br>5 |
| | B | **5**<br>80 | **20**<br>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.

---

[20] Appendices B, C and D are not intended for publication.

[21] 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!

### INSTRUCTIONS FOR THE PARTICIPANTS IN TREATMENT 3 AND 4 IN STUDY 1

**Stock Market Game [Community Game][22], 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.

[22] 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 com-puter / the receiver | A | **50**<br>50 | **80**<br>5 |
| | B | **5**<br>80 | **20**<br>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 C: INSTRUCTIONS FOR THE PARTICIPANTS IN STUDY 2

## STOCK MARKET GAME [COMMUNITY GAME][23], 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**<br>A | <br>B |
|---|---|---|---|
| The computer / the receiver | A | **50**<br>50 | **80**<br>5 |
| | B | **5**<br>80 | **20**<br>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.

[23] 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 D: INSTRUCTIONS FOR THE PARTICIPANTS IN STUDY 3

## STOCK MARKET GAME [COMMUNITY GAME][24], 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 |
| *Player 2* | A | **50**<br>50 | **80**<br>5 |
| | B | **5**<br>80 | **20**<br>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.

[24] 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!