Games and Economic Behavior 76 (2012) 117–130

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Social framing effects: Preferences or beliefs?

Tore Ellingsen [a,*], Magnus Johannesson [a], Johanna Mollerstrom [b], Sara Munkhammar [c]

[a] Stockholm School of Economics, Box 6501, 13183 Stockholm, Sweden
[b] Harvard University, Department of Economics, Littauer Center 200, Cambridge, MA 02138, USA
[c] Trygg-Hansa, Fleminggatan 18, 106 26 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 20 April 2011
Available online 17 May 2012

JEL classification:
D03
A13
C92

Keywords:
Framing
Labels
Social preferences

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 her action. The social framing effect also vanishes when the game is played sequentially. These findings are inconsistent with the hypothesis that the Community label triggers a desire to cooperate, but consistent with the hypothesis that social frames are coordination devices. More generally, our evidence indicates that social frames enter people's beliefs rather than their preferences.

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 also affect behavior. 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 et al., 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 et al., 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

* Corresponding author.
E-mail addresses: tore.ellingsen@hhs.se (T. Ellingsen), magnus.johannesson@hhs.se (M. Johannesson), jmollers@fas.harvard.edu (J. Mollerstrom).

[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), Brandts and Schwieren (2009), Brewer and Kramer (1986), Cookson (2000), Cubitt et al. (2011), Dufwenberg et al. (2011), McDaniel and Sistrunk (1991), McCusker and Carnevale (1995), Meier (2006), Pillutla and Chen (1999), Rege and Telle (2004), Sell and Son (1997), van Dijk and Wilke (2000), and Zhong et al. (2007). Labels have also been shown to affect cooperative behavior in other games; see for example Larrick and Blount (1997) and Barr and Serra (2009). 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 et al. (1998) for a survey of individual choice effects of wording.

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2012.05.007

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, 2001), and other prosocial behaviors (Koneberg et al., 2010). Through this channel, the very language and assumptions of economics could be eroding cooperation (Ferraro et al., 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, p. 75) 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 at least three broad classes of social framing theories. The first class posits that frames affect internalized social norms or, alternatively, social preferences[3]: This would mean that the Community label triggers a stronger desire or compulsion to cooperate (Montgomery, 1998; Weber et al., 2004). We call this the *variable sociality hypothesis*.

Bacharach (2006) provides a formal treatment of the variable sociality hypothesis for a particular case of internalized social norms, namely team reasoning.[4] An alternative formalization, involving a smaller departure from conventional game theory, is to assume that certain decision frames affect social preferences, such as altruism. In this case, a Prisoners' dilemma in material payoffs may be transformed into, for example, a common interest game 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 not.[5]

Another hypothesis is that people respond to social frames because the frame affects how others interpret their behavior, which in turn determines their social esteem. Even a person who has not internalized norms, and who holds strictly selfish preferences, may want to appear to be prosocial. This is the *social image hypothesis*.[6]

A third class of theories of social framing effects is that the frame affects the expectations that people have about each other's behavior, and these expectations in turn affect the own behavior. For this theory to apply in a Prisoners' dilemma, it is necessary 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. 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 outset, we lump together all these models that are based on multiple equilibria and refer to them as the *coordination hypothesis*.

In order to evaluate the relative importance of the three hypotheses, we first state them formally. We then report three separate one-shot Prisoners' dilemma experiments, in which we systematically vary features of the game. The first experiment replicates the main finding of previous studies, namely the presence of framing effects in a simultaneous Prisoners' dilemma. It also shows that this finding is difficult to explain with some versions of the variable sociality hypothesis. The second experiment evaluates the social image hypothesis, which finds no support in the data. The third experiment considers framing effects in sequential Prisoners' dilemmas. It finds none.

In our view, the 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.[7] This finding is consistent with the coordination hypothesis, but inconsistent with the variable sociality 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, team reasoning or altruism 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 modern version of the rational choice paradigm, without any appeal to context-dependent preferences. (As we shall see, one caveat is that the beliefs themselves could be involving some notion that opponents have – or expect their opponents to have etc. – context dependent preferences.)

The paper is most closely related to Liberman et al. (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.[8] The second study instead uses 40 Israeli pilot trainees, and the labels Bursa Game and Kommuna Game, but is

[3] Social norms are defined on the set of social situations (i.e., game forms; see below), whereas social preferences are typically defined on some set of ultimate outcomes.

[4] Roughly, when a person engages in team reasoning, she asks which strategy profile is best for the group and picks her component of that profile; see Sugden (1993, 2003) and Bacharach (1999, 2006). The theory comes in several flavors; Bacharach considers unconditional norm compliance, whereas Sugden (2003) emphasizes that compliance may be conditional on the belief that others comply as well.

[5] The term "game form" originates from Gibbard (1973, p. 587), 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).

[6] 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.

[7] It would have been even better to have data from one experiment in which framing effects under simultaneous and sequential play is compared directly. We return to this and other caveats in the two final sections.

[8] 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.

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

Fig. 1. Material payoffs – the game form.

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.[9] 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[10]: 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 et al. (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. (Of course, even if they are not affected by the frame, preferences are still important in our analysis, since they determine whether beliefs matter for behavior to begin with.)

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.[11]

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.[12] Although Yamagishi and Kiyonari do not explicitly invoke the game theoretic argument, it is clear that their idea is similar to ours: They interpret the sharp reduction of in-group favoritism in the sequential setting as an indication 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.

Our findings also relate quite closely to Bohnet and Cooter (2001), 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 a suitably 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 discusses some caveats and concludes.

2. Theory

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

[9] If the game had been one-shot, the latter difference would clearly have suggested that preferences depend on frames. However, here it might instead reflect different expectations about behavior in later rounds.

[10] Cubitt et al. (2011) provide further reasons for studying framing effects in one-shot games.

[11] Note however that two recent studies, Brandts and Schwieren (2009) and Dufwenberg et al. (2011), 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.

[12] 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.

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

Fig. 1a. The game played by frame-sensitive altruists.

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 Fig. 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 $\{C, D\}$ be a pure action for player i and let $m_i(S_i, S_j)$ 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, or they may have a view about what is the right course of action in a certain situation, that is, they may care about social norms.

Players may also care about non-material outcomes, such as the opponent's belief about their social preferences. To capture such preferences, let $\mathfrak{I}$ 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 ${}^e\tau_{ji}$ denote player i's expectation about τ_{ji}.[13] Finally, let $\mathfrak{F}$ denote the set of frames, with typical element F. In general, player i's utility can thus be written as a function $U_i(m_i, m_j, \tau, \tau_{ij}, {}^e\tau_{ji}, F)$.

Below, we consider some specific formal examples of how the frame may enter players' preferences and beliefs respectively. In most of the examples, we abstract from preference heterogeneity in order to simplify the analysis. With heterogeneous preferences, we would have to employ different solution concepts, using the theory of games with incomplete information. A major benefit from taking that step is that we would be better able to fit the within-treatment variation in the experimental data. In the fifth model considered below, we admit preference heterogeneity and show that the key arguments concerning across-treatment variation are the same as without preference heterogeneity.

2.1. Frame-dependent preferences

The frame could be directly affecting the degree of unselfishness. Suppose 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

$$U_i = m_i + \alpha(F)m_j. \tag{1}$$

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

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, we have that the game is Stag hunt if $(d-s)/(w-d) > (w-c)/(c-s)$, and Chicken if $(d-s)/(w-d) < (w-c)/(c-s)$.)

Team reasoning, at least in the sense of Bacharach (2006), can be seen as an internalized social norm that requires players to unconditionally pick the strategy profile that is consistent with joint payoff maximization, here (C, C).[14] For the purposes of our study, team reasoning thus yields the same behavior as an altruism parameter $\alpha > \max\{(w-c)/(c-s), (d-s)/(w-d)\}$.[15] Indeed, perhaps the most natural interpretation of the players' altruistic concern $\alpha(F)$ is that they have internalized an efficiency norm, not that they care strongly about the opponent's life-time consumption.[16] 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.

[13] 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 et al. (2006), Broberg et al. (2007), and Lazear et al. (2012). Beliefs may affect preferences also in a model without "types" – see Dufwenberg et al. (2011) for references and an application to framing effects.

[14] Sugden (2003) proposes a theory of team reasoning in which players only want to stick to the norm if they expect opponents to do so too. This model of expectation formation entails coordination on efficient equilibria in common interest games.

[15] As noted by Sugden, 2008, there is always such correspondence between team reasoning and linear social preferences when the game is decomposable; however, for non-decomposable games the correspondence breaks down.

[16] See, Andreoni and Bernheim (2009), Krupka and Weber (2010), and especially López-Pérez (2008) for different models of norm compliance.

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

Fig. 1b. The game with altruism towards a worse off opponent.

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. For an economist, this is perhaps the most straightforward formalization of the variable sociality hypothesis.

The social esteem hypothesis instead 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

$$U_i = m_i + \alpha(F)m_j + v\big({}^e\alpha_{ji}(S_i, F), F\big). \tag{2}$$

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. Note here that the opponent's ex post belief, and hence one's esteem, may depend on the own action. In general, the frame could affect both the opponent's interpretation (and hence be an argument of ${}^e\alpha_{ji}$) and the utility of the opponent's belief – it's more valuable to be considered unselfish in a community setting (and hence appear as an independent argument of v).

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, $v({}^e\alpha_{ji}, F)$, is greater under one frame than another, then this hypothesis works in essentially the same way as the variable sociality hypothesis.[17] 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 $m_i \geqslant m_j$ and 0 otherwise.[18] Then, if preferences are common knowledge, the game is as depicted in Fig. 1b.

If $\alpha > (w-c)/(c-s)$, the game has two pure strategy equilibria, namely (C, C) and (D, D), where the former equilibrium Pareto-dominates the latter. That is, the game is not Prisoners' dilemma, but Stag hunt.[19]

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

$$U_i = m_i - \beta \max\{0, m_i - m_j\}. \tag{4}$$

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

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

There are several other versions 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

[17] 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.)

[18] 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.

[19] We here employ a broad definition of Stag hunt. A narrower definition would admit only games in which (D, D) is risk-dominant. In our example, this condition is fulfilled if s is sufficiently small.

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

Fig. 1c. The game when players are averse to advantageous inequality.

and Rabin (2002), here is one model that we like particularly well, and which will also turn out to rationalize our data. Suppose there are two types of players, egoists and conditional altruists with the utility function

$$U_i = m_i + \alpha I \alpha_{ij} m_j, \tag{5}$$

where I is an indicator variable taking the value 1 if $m_i \geqslant m_j$ 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 see that the frame can be used as a coordination device, indicating a 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.[20]

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 et al. (1994) and Crawford et al. (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). [21]

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 a level-1 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, will also cooperate under the Community label but not under the Stock Market label.

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 at the 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 SEK[22] (Swedish Kronor; \$1 $\approx$ 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 Fig. 2.

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

[21] Cachon and Camerer (1996) and Rydval and Ortmann (2005) demonstrate that loss aversion furnishes another, and possibly related, coordination principle.

[22] 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

Fig. 2. The game form in the experiment.

Since each subject earns more by choosing B than by choosing A, and 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 Fig. 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 the current text.

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.[23] 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 behavior is the same in T3 as in T1, and the same in T4 as in T2. Thus, the framing effect ought to be 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.[24] 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

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

which simplifies to

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

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

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

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

[23] The relevant part of the Instructions read: "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."

[24] McCabe et al. (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.

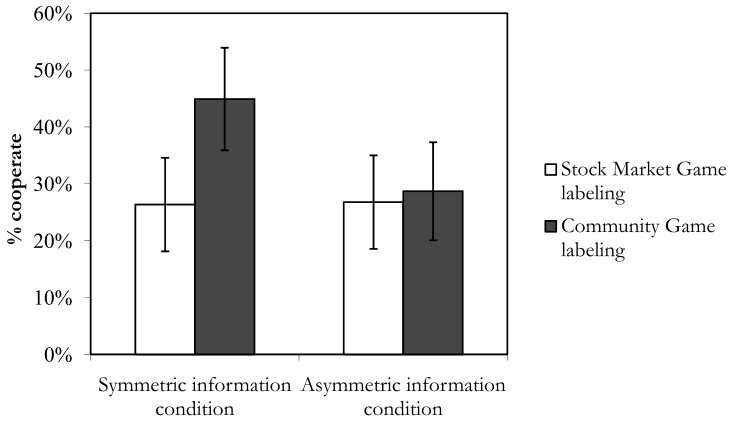

Fig. 3. Fraction of cooperative actions in treatments 1–4. The difference between the first two bars is the social framing effect in the standard symmetric information condition. The difference between the second two bars is the social framing effect in the asymmetric information condition. Error bars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

which simplifies to

$$\mu\big((13\mu - 4)\alpha - 3\big) \geqslant 3.$$

Again, we can use the fact that μ lies between 0 and 1 to deduce the necessary condition $\alpha \geqslant 2/3$, just as before. However, the additional requirement on the two parameters is stronger than before. (To see this, rewrite the condition as $\alpha\mu(13\mu - 4) \geqslant 3(1 + \mu)$. This inequality is never satisfied if $\mu \leqslant 4/13$. If instead $\mu \geqslant 4/13$, the condition can be written $\alpha \geqslant 3(1 + \mu)/(13\mu - 4)\mu$, which is implied by the previous condition $\alpha \geqslant (1 + \mu)/3\mu$ 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 Fig. 3.[25] 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; $z = 2.972$, $p = 0.0015$),[26] rejecting the null hypothesis of zero framing effect.

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% 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.[27]

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 variable sociality 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,

[25] 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 bars of Fig. 7.

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

[27] 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 ($z = 0$, $p = 0.5$).

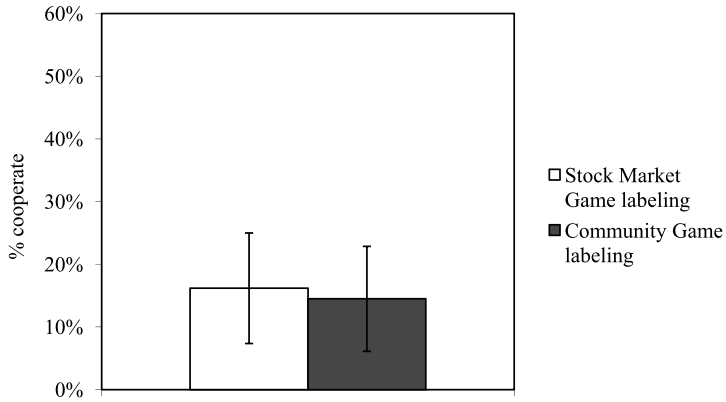

Fig. 4. Fraction of cooperative actions in the second experiment, treatments 1′ and 2′. The difference between the two bars is the social framing effect in the symmetric information passive opponent condition. Error bars indicate 95% confidence intervals.

hold other beliefs in the active condition than in the passive condition. Liberman et al. (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 et al., 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).[28]

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, as is admitted by Model (2), framing might matter more in treatments 1 and 2 than in 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 26, but the baseline level of cooperation is lower. Therefore, any difference in framing effects is unlikely to be caused by subject pool effects.

Fig. 4 displays the findings.

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

[28] We are grateful to Ernst Fehr for this suggestion.

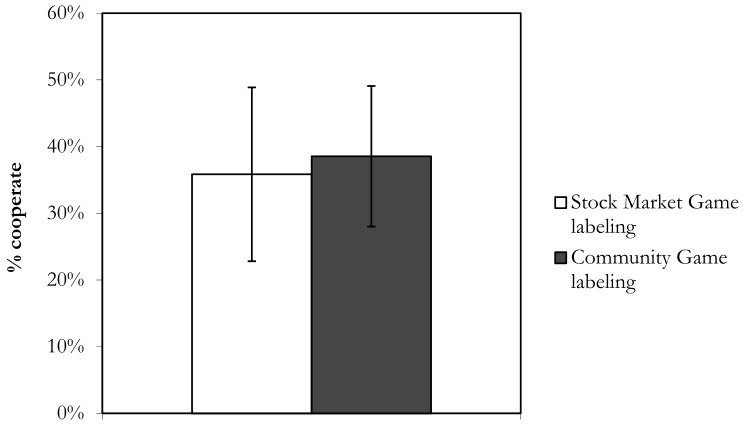

Fig. 5. Fraction of first-mover cooperative actions in the third experiment treatments 1″ and 2″. The difference between the two bars is the social framing effect for player 1 in the sequential moves condition. Error bars indicate 95% confidence intervals.

condition, we find no support for the hypothesis that the social framing effect in Study 1 was caused by social esteem considerations. That is, the evidence is unsupportive of 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 variable sociality hypothesis and the coordination hypothesis, namely by letting the moves be sequential instead of simultaneous.[29] 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.[30] On the other hand, if the variable sociality 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 study 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 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.

Fig. 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.[31] 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 variable sociality hypothesis.

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 variable sociality hypothesis. Fig. 6 displays the results. The first pair of bars 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 bars gives the corresponding cooperation rates for the case in which player 1 defects.

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, $z = 0.511$, $p = 0.305$).

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

[30] 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 et al. (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.

[31] Recall that the high average cooperation rates in T1 and T2 is driven by the Södertörn subjects. See also footnote 26.

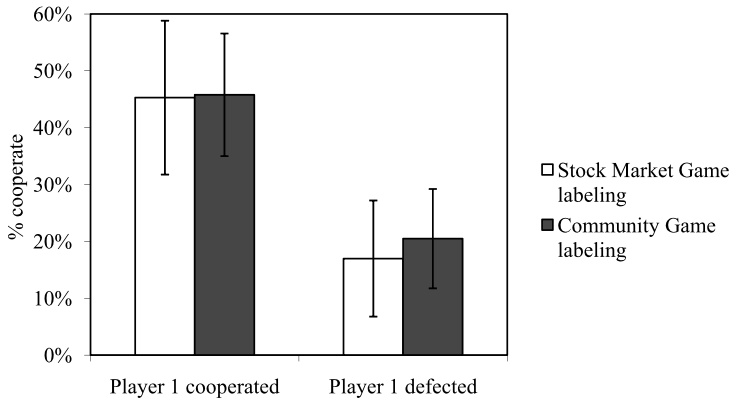

Fig. 6. Fraction of second-mover cooperative actions in the third experiment treatments 1″ and 2″. The difference between the first (second) two bars is the social framing effect for player 2, when player 1 cooperated (defected), in the sequential moves condition. Error bars indicate 95% confidence intervals.

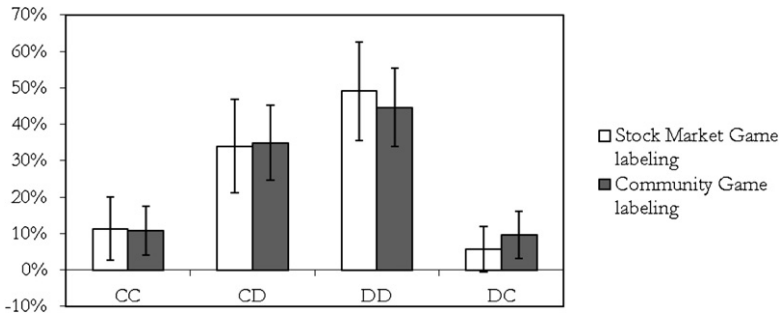

Fig. 7. Distribution of second-mover strategies in the third experiment treatments 1″ and 2″. Each staple indicates the fraction of second movers that chose the strategy in the Stock Market frame (first staple in each pair) and the Community frame (second staple) respectively. Error bars indicate 95%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

The lack of significant differences is further emphasized in Fig. 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.[32] 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 $= 0.789$, $p = 0.852$). Among players who cooperate after cooperation, there is no difference at all. Among players who defect after cooperation, cooperation after defection is somewhat more frequent in the Community treatment, but the effect is statistically insignificant ($p = 0.192$).

In our view, the last experiment contradicts 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 "variable sociality" is one.[33] 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 variable altruism 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.

We also acknowledge the possibility that the subject pool in 2009 was significantly different from the pool in 2006. While the measurable SSE student characteristics are not likely to have changed (the program is exceptionally popular and

[32] 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.

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

therefore attracts the top new students every year), it is quite possible that important events such as the financial crisis may have affected the students' sensitivity to social frames. If students have become less prone to behave selfishly under the stock market frame, for example due to the public criticism of greedy bankers, that might have eliminated frame-sensitive sociality in 2009 even if it was there in 2006.

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 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, whose analysis often rest on the assumption that preferences are stable and reasonably simple, a methodological principle most stoutly defended by Becker and Stigler (1977). Conversely, the results suggest that some of the critics of economic theory, as exemplified by Ferraro et al. (2005) and references therein, might shift their focus from economists' modeling of preferences to their modeling of beliefs.

Among the five explicit models that we consider, a version of the social preferences proposed by Charness and Rabin (2002) is the only model not to be severely questioned 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 adapted to articulate more precisely how framing affect coordination. However, such a complete non-equilibrium analysis is beyond the scope of this paper.

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 variable sociality hypothesis, since our preferred model leaves no room for coordination in a game with only a single active player.[34] However, in support of the present analysis, the data indicate no social framing effects in the Dictator game.

Although our study is large by current standards, our findings call for replication. Since inter-experiment comparisons are always less credible than intra-experiment comparisons, we would especially like to see a single study that directly compares the impact of social frames in simultaneous and sequential Prisoners' dilemmas.

Acknowledgments

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, many seminar participants, and especially several anonymous referees, for helpful comments. Thanks 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.

Supplementary material

The online version of this article contains additional supplementary material.
Please visit http://dx.doi.org/10.1016/j.geb.2012.05.007.

References

Andreoni, J., 1995. Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. Q. J. Econ. 110, 1–21.
Andreoni, J., 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: Binmore, K., Kirmanand, A., Tani, P. (Eds.), Frontiers of Game Theory. MIT Press, Cambridge, MA, pp. 255–276.
Bacharach, M., 1999. Interactive team reasoning: A contribution to the theory of co-operation. Res. Econ. 53, 117–147.
Bacharach, M., 2006. In: Gold, N., Sugden, R. (Eds.), Beyond Individual Choice. Princeton University Press, Princeton.
Bacharach, M., Bernasconi, M., 1997. The variable frame theory of focal points: An experimental study. Games Econ. Behav. 19, 1–45.
Bacharach, M., Stahl, D.O., 2000. Variable frame level-n theory. Games Econ. Behav. 33, 220–246.
Bardsley, N., Mehta, J., Starmer, C., Sugden, R., 2010. Explaining focal points: Cognitive hierarchy theory versus team reasoning. Econ. J. 120, 40–79.

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

Barr, A., Serra, D., 2009. The effects of externalities and framing on bribery in a petty corruption experiment. Exp. Econ. 12, 488–503.
Batson, C.D., Moran, T., 1999. Empathy-induced altruism in a Prisoner's dilemma. Eur. J. Soc. Psychol. 29, 909–924.
Becker, G., 1974. A theory of social interactions. J. Polit. Econ. 82, 1063–1093.
Becker, G.S., Stigler, G.J., 1977. De gustibus non est disputandum. Am. Econ. Rev. P&P. 67, 76–90.
Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96, 1652–1678.
Bernheim, B.D., 1994. A theory of conformity. J. Polit. Econ. 102, 841–877.
Bohnet, I., Cooter, R.D., 2001. Expressive law: Framing or equilibrium selection. Mimeo, Harvard University.
Brandts, J., Schwieren, C., 2009. Frames and economic behavior: An experimental study. Mimeo, Universitat Autònoma de Barcelona.
Brennan, G., Pettit, P., 2004. The Economy of Esteem. Oxford University Press, Oxford.
Brewer, M.B., Kramer, R.M., 1986. Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. J. Pers. Soc. Psychol. 50, 543–549.
Broberg, T., Ellingsen, T., Johannesson, M., 2007. Is generosity involuntary? Econ. Lett. 94, 32–37.
Burnham, T., McCabe, K.A., Smith, V.L., 2000. Friend-or-foe intentionality priming in an extensive form trust game. J. Econ. Behav. Organ. 43, 57–73.
Cachon, G.P., Camerer, C.F., 1996. Loss-avoidance and forward induction in experimental coordination games. Q. J. Econ. 111, 165–194.
Camerer, C.F., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, New Jersey.
Casari, M., Cason, T.N., 2009. The strategy method lowers measured trustworthy behavior. Econ. Lett. 103, 157–159.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Q. J. Econ. 117, 817–869.
Chen, Y., Li, S.X., 2009. Group identity and social preferences. Am. Econ. Rev. 99, 431–457.
Clark, K., Sefton, M., 2001. The sequential Prisoners' dilemma: Evidence on reciprocation. Econ. J. 111, 51–68.
Cookson, R., 2000. Framing effects in public goods experiments. Exp. Econ. 3, 55–79.
Crawford, V.P., Gneezy, U., Rottenstreich, Y., 2008. The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures. Am. Econ. Rev. 98, 1443–1458.
Croson, R.T.A., 1999. The disjunction effect and reason-based choice in games. Organ. Behav. Hum. Decis. Process. 80, 118–133.
Costa-Gomes, M.A., Weizsäcker, G., 2008. Stated beliefs and play in normal-form games. Rev. Econ. Stud. 75, 729–762.
Cubitt, R.P., Drouvelis, M., Gächter, S., 2011. Framing and free riding: Emotional responses and punishment in social dilemma games. Exp. Econ. 14, 254–272.
Dana, J., Cain, D.M., Dawes, R.M., 2006. What you don't know won't hurt me: Costly but quiet exit in dictator games. Organ. Behav. Hum. Decis. Process. 100, 193–201.
Dawes, R.M., McTavish, J., Shaklee, H., 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. J. Pers. Soc. Psychol. 35, 1–11.
Deutsch, M., 1958. Trust and suspicion. J. Conflict Resolut. 2, 265–279.
Deutsch, M., 1960. The effect of motivational disposition on trust and suspicion. Hum. Rel. 13, 123–139.
Dufwenberg, M., Gächter, S., Hennig-Schmidt, H., 2011. The framing of games and the psychology of play. Games Econ. Behav. 73, 459–478.
Eiser, J.R., Bhavnani, K., 1974. The effect of situational meaning on the behaviour of subjects in the Prisoner's Dilemma Game. Eur. J. Soc. Psychol. 4, 93–97.
Ellingsen, T., Johannesson, M., 2007. Paying respect. J. Econ. Perspect. 21 (4), 135–149.
Ellingsen, T., Johannesson, M., 2008. Pride and prejudice: The human side of incentive theory. Am. Econ. Rev. 98, 990–1008.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition and cooperation. Q. J. Econ. 114, 817–868.
Fehr, E., Schmidt, K., 2006. The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In: Kolm, S.C., Ythier, J.M. (Eds.), Handbook of the Economics of Giving, Altruism and Reciprocity, vol. 1. Elsevier, Amsterdam.
Ferraro, F., Pfeffer, J., Sutton, R.I., 2005. Economics language and assumptions: How theories can become self-fulfilling. Acad. Manage. Rev. 30, 8–24.
Frey, B.S., Osterloh, M., 2005. Yes managers should be paid like bureaucrats. J. Manage. Inq. 14, 96–111.
Gibbard, A., 1973. Manipulation of voting schemes: A general result. Econometrica 41, 587–601.
Glazer, A., Konrad, K.A., 1996. A signaling explanation of charity. Am. Econ. Rev. 86, 1019–1028.
Ireland, N., 1994. On limiting the market for status signals. J. Public Econ. 53, 91–110.
Kay, A.C., Ross, L., 2003. The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner's dilemma. J. Exp. Soc. Psychol. 39, 634–643.
Koneberg, C., Yaish, M., 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. Rat. Society 22, 3–36.
Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated Prisoners' dilemma. J. Econ. Theory 27, 245–252.
Krupka, E., Weber, R.A., 2010. Identifying social norms using coordination games: Why does dictator game sharing vary? Mimeo, Carnegie Mellon University.
Larrick, R.P., Blount, S., 1997. The claiming effect: Why players are more generous in social dilemmas than in Ultimatum games. J. Pers. Soc. Psychol. 72, 810–825.
Lazear, E.P., Malmendier, U., Weber, R.A., 2012. Sorting in experiments with application to social preferences. Am. Econ. J. Appl. Econ. 4, 136–163.
Levin, I.P., Schneider, S.L., Gaeth, G.J., 1998. All frames are not created equal: A typology and critical analysis of framing effects. Organ. Behav. Hum. Decis. Process. 76, 149–188.
Levitt, S., List, J., 2007. What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. 21 (2), 151–174.
Liberman, V., Samuels, S.M., Ross, L., 2004. The name of the game: Predictive power of reputations versus situational labels in determining Prisoner's dilemma game moves. Pers. Soc. Psychol. Bull. 30, 1175–1185.
López-Pérez, R., 2008. Aversion to norm-breaking: A model. Games Econ. Behav. 64, 237–267.
McCabe, K.A., Rigdon, M.L., Smith, V.L., 2003. Positive intentions and intentions in trust games. J. Econ. Behav. Organ. 52, 267–275.
McCusker, C., Carnevale, P.J., 1995. Framing in resource dilemmas: Loss aversion and the moderating effects on sanctions. Organ. Behav. Hum. Decis. Process. 61, 190–201.
McDaniel, W.C., Sistrunk, F., 1991. Management dilemmas and decisions: impact of framing and anticipated responses. J. Conflict Resolut. 35, 21–42.
Mehta, J., Starmer, C., Sugden, R., 1994. The nature of salience: An experimental investigation. Am. Econ. Rev. 84, 658–673.
Meier, S., 2006. Does framing matter for conditional cooperation? Evidence from a natural field experiment. Contrib. Econ. Anal. Policy 5 (2). Article 1.
Montgomery, J.D., 1998. Toward a role-theoretic conception of embeddedness. Am. J. Sociol. 104, 92–125.
Pfeffer, J., 2007. Human resources from an organizational behavior perspective: Some paradoxes explained. J. Econ. Perspect. 21 (4), 115–134.
Pillutla, M.M., Chen, X., 1999. Social norms and cooperation in social dilemmas: The effects of context and feedback. Organ. Behav. Hum. Decis. Process. 78, 81–103.
Rabin, M., 1993. Incorporating fairness into game theory and economics. Am. Econ. Rev. 83, 1281–1302.
Rabin, M., 1998. Psychology and economics. J. Econ. Lit. 36, 11–46.
Radner, R., 1986. Can bounded rationality resolve the Prisoners' dilemma?. In: Hildenbrand, W., Mas-Colell, A. (Eds.), Contributions to Mathematical Economics. North-Holland, Amsterdam.
Rege, M., Telle, K., 2004. The impact of social approval and framing on cooperation in public good situations. J. Public Econ. 88, 1625–1644.
Rydval, O., Ortmann, A., 2005. Loss avoidance as selection principle: Evidence from simple stag-hunt games. Econ. Lett. 88, 101–107.

Schelling, T.C., 1960. The Strategy of Conflict. Harvard University Press, Cambridge, MA.
Sell, J., Son, Y., 1997. Comparing public goods and common pool resources: three experiments. Soc. Psychol. Q. 60, 118–137.
Selten, R., 1967. Die Strategiemethode zur Erforschung des eingeschränkt rationalen. In: Sauermann, H. (Ed.), Verhaltens im Rahmen eines Oligopolexperiments. In: Beiträge zur experimentellen Wirtschaftsforschung, vol. I. J.C.B. Mohr (Siebeck), Tübingen, pp. 136–168.
Sen, A.K., 1967. Isolation, assurance, and the social rate of discount. Q. J. Econ. 81, 112–124.
Sobel, J., 2005. Interdependent preferences and reciprocity. J. Econ. Lit. 43, 392–436.
Sugden, R., 1993. Thinking as a team: Towards an explanation of nonselfish behavior. Soc. Philos. Policy 10, 69–89.
Sugden, R., 2003. The logic of team reasoning. Philos. Expl. 6, 165–181.
Sugden, R., 2008. Nash equilibrium, team reasoning and cognitive hierarchy theory. Acta Psychol. 128, 402–404.
Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211, 453–458.
Tyran, J.-R., Feld, L., 2006. Achieving compliance when legal sanctions are non-deterrent. Scan. J. Econ. 108, 135–156.
van Dijk, E., Wilke, H., 2000. Decision-induced focusing in social dilemmas: Give-some, keep-some, take-some, and leave-some dilemmas. J. Pers. Soc. Psychol. 78, 92–104.
Weber, J.M., Kopelman, S., Mssick, D.M., 2004. A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. Pers. Soc. Psychol. Rev. 8, 281–307.
Yamagishi, T., Kiyonari, T., 2000. The group as the container of generalized reciprocity. Soc. Psychol. Q. 63, 111–132.
Zhong, C.-B., Loewenstein, J., Murnighan, J.K., 2007. Speaking the same language: The cooperative effects of μ in the Prisoners' dilemma. J. Conflict Resolut. 51, 431–456.