Kamei, Kenju; Putterman, Louis

Working Paper

Reputation transmission without benefit to the reporter: A behavioral underpinning of markets in experimental focus

Working Paper, No. 2015-9

Provided in Cooperation with:
Department of Economics, Brown University

Suggested Citation: Kamei, Kenju; Putterman, Louis (2015) : Reputation transmission without benefit to the reporter: A behavioral underpinning of markets in experimental focus, Working Paper, No. 2015-9, Brown University, Department of Economics, Providence, RI

This Version is available at:
https://hdl.handle.net/10419/145432

Reputation Transmission without Benefit to the Reporter:

a Behavioral Underpinning of Markets in Experimental Focus

Kenju Kamei and Louis Putterman*

This version: September 8, 2015

Abstract:

Reputation is a commonly cited check on opportunism in economic and social interactions. But it is often unclear what would motivate an agent to report another's behavior when the pool of potential partners is large and it is easy enough for an aggrieved player to move on. We argue that behavioral or social preference motivations may solve this conundrum. In a laboratory experiment in which subjects lack any private material incentive to report partners' actions, we find that most cooperators incur a cost to report a defecting partner when this has the potential to deprive the latter of future gains and to help his next partner.

JEL classification: C91, D03, D63

Keywords: reputation, prisoners' dilemma, experiment, punishment, communication, costly reporting, social preference, inequity aversion

* Department of Economics and Finance, Durham University (email: kenju.kamei@durham.ac.uk, kenju.kamei@gmail.com) and Department of Economics, Brown University (email: Louis_Putterman@brown.edu), respectively. Funding for this project came from the Murata Science Foundation, with additional support from Brown University. We wish to thank Cheewadhanaraks Matanaporn for programming of the experiment software, and Benjamin Chowdhury, Kathleen Hu and especially Hui Wen Ng for assistance with the literature review, theoretical analysis and data analysis. We also thank Yan Chen for her hospitality while the authors conducted the experiments at the University of Michigan, and we thank Pedro Dal Bó and other participants at Brown University theory lunch, at the 2015 Social Dilemmas Conference at Brown, and at the ESA World Meeting in Sydney, Australia in July 2015 for helpful comments.

1. Introduction

People face many situations in which there exist potential benefits of cooperating with others and accompanying dangers of being exploited by them. The fact that it is sometimes possible to choose interaction partners based on their reputations for cooperativeness and that one might accordingly have an incentive to invest in such a reputation, thus incentivizing cooperation where defection would otherwise be selfishly rational, is an oft-noted factor supporting cooperation. But what motivates actors to transmit reputational information? What, exactly, is the motivating force that underlies the relevant information's transmission? We conjecture that the same factors which lead to costly punishment of unfair actors in social dilemma experiments (Ostrom *et al*., 1992; Fehr and Gächter, 2000; Gächter and Herrmann, 2009) can lead to costly reporting of opportunistic interaction partners when there is no potential material gain to the individual doing the reporting.

Societies invest considerable resources in workarounds of dilemmas of cooperation. People are taught to cooperate with members of their own group, family, or nation, are told that they will be held in others' contempt if caught cheating, and are taught to believe in supernatural reward or punishment for doing so or failing to. That such investments in inculcating trustworthiness are ubiquitous hints that at least some people are receptive to them (Wilson, 2002). Such receptivity is compatible with the outlooks of both sentiment-focused moral philosophers like Adam Smith (1761) and of more recent evolutionary psychologists who argue that nature (evolution) itself tinkered with our psyches during the hundreds of millennia of small band existence to facilitate cooperation (Sobel, 2005; Bowles and Gintis, 2011). Built-in inclinations that help us to overcome dilemmas of cooperation may include hyper-sensitivity to and special interest in signs of others' cheating, emotions of anger supporting a willingness to

expend resources to punish cheaters, and emotions of guilt and shame that may function as self-punishments even when our indiscretions are not detected. The fact that people will sometimes discharge obligations to risk death on a battlefield or to charge into a burning building in part because living with the shame of not doing so seems worse than death, speaks to the power of these elements of our psychological natures (Field, 2003; Gintis *et al*., 2005; Wilson, 2012).

One of the phenomena to which evolutionary psychologists point as evidence of near-universal psychological adaptations to the challenges posed by social dilemma problems is the observation that gossip is a major component of social interaction in every known culture (Pinker, 2003; Dunbar, 2004). A considerable fraction of gossip consists of reports of others' misdeeds, or of reasons for doubting their adherence to appropriate norms of behavior. Feinberg *et al*. (2012) report evidence that "individuals who observe an antisocial act experience negative affect and are compelled to share information about the antisocial actor with a potentially vulnerable person," and that "individuals possessing more prosocial orientations are the most motivated to engage in such gossip, even at a personal cost" (p. 1015). Such information appears to have a distinct power to grab attention, and psychological normalcy includes a desire to avoid being the subject of negative gossip.[1]

Aside from its psychic cost, being the subject of negative reports may also have its direct material consequences, if others hesitate to engage in cooperative interactions with individuals known for past reneging or defection. This often can be the case under modern circumstances, in which individuals have multiple choices with respect to where to shop, which plumber to hire, and so forth. But the same mobility that puts potential cooperation partners in competition with each other may pose a threat to the reputational mechanism through which competition operates,

[1] As Adam Smith (1761) wrote: "Man naturally … dreads, not only blame, but blame-worthiness."

if it is costly to spread the word about an interaction partner's defection and if one has already decided to leave that partner behind. How will others learn whom to avoid, if the victims of unscrupulous agents' actions lack the incentive to report on them?

If all victims are the actors of traditional economic theory who care only about own material payoffs, if they cannot profitably exchange the information, and if there is any cost associated with conveying it, then there will be no such reporting.[2] However, possible motivations for the costly reporting of cheating or opportunistic behavior may be identified when factors additional to material payoff are taken into consideration. First, engaging in negative gossip may be a direct source of satisfaction built into the human psyche, paralleling reports that pleasure centers in the brain "light up" when experimental subjects punish selfish counterparts in a trust game (de Quervain *et al*., 2004). Second, tipping others off not to naively cooperate with the miscreant may bring satisfaction not for the act itself but thanks to anticipation of the punishment that this may visit on that actor. The large literature showing cooperative subjects' willingness to spend money to reduce free riders' earnings in voluntary contribution experiments (Falk *et al.*, 2005; Gächter and Herrmann, 2009; Putterman, 2014) suggests that such motives are widespread, and raises the possibility that punitive motives might also motivate *reporting as punishment*. Third, conceivably the victim feels empathy or obligation towards others who are in danger of being victimized, and may accordingly try to warn them. While all of these factors may involve emotional responses of one kind or another, the second and third factors, at least, may be amenable to representation as social preferences that figure in the utility-maximizing calculations of strictly rational (but not perfectly selfish) agents.

[2] Note that in network theory (Jackson and Zenou, 2014), the standard assumption is that while creation of links may itself be costly, information travels costlessly among those actors who are linked together, and that information is passed along by default whenever two agents are linked and the informed agent is indifferent to having transmission occur. We could locate no discussion of behavioral or social preference explanations of willingness to incur costs to transmit information, in this literature.

Given the importance of reputational mechanisms to solving dilemmas of cooperation and the rapid growth of behavioral economics and social preference research using the techniques of experimental economics, one might expect the costly reporting of partners' behaviors to be the topic of numerous studies. Yet we were unable to locate such studies in the literature.[3] We address this gap by conducting experiments with which to investigate both the willingness to pay to report uncooperativeness, and subjects' beliefs about how common that willingness is. Our experimental design, which builds on familiar prisoners' dilemma stage games, allows us to isolate the willingness to engage in reporting when a personal monetary payoff is unavailable, so that only desires to punish the individual or to protect her future interaction partner are potential motives for reporting. We demonstrate that reporting is significantly less common when it is costly than when free, but that costly reporting does occur often, with cooperators-defector reporting being far more common than cooperator-cooperator, defector-cooperator, or defector-defector reports. Indeed, a majority of cooperators meeting defectors report them despite its cost and lack of personal benefit. We identify conditions under which such reporting is consistent with a familiar model of social preferences, the inequity aversion model of Fehr and Schmidt (1999).

Although reporting at positive cost is never rational for strictly selfish agents in our setting, the same does not apply to cooperating. We conduct an incentivized elicitation of beliefs about others' cooperation and reporting, and with the resulting data we calculate which subjects could be rationally choosing or rejecting cooperation out of simple payoff maximization, and which choose cooperation (or defection) in error or due to a social preference or emotion.

[3] There are studies of leaving reviews online, a few of which we cite in the next section, but none that we could find systematically focuses on the costliness of reviewing, nor is it easy to see how studies of reporting in that setting could achieve the degree of control we obtain in the lab.

Almost all observed decisions to defect in the experiment are explicable by payoff maximization under own beliefs regarding others' cooperation and reporting probabilities. Rational maximization of own payoff also explains many choices to cooperate, but a substantial number of cooperation choices require alternative explanation, with the Fehr-Schmidt model also a workable candidate in many cases. In addition to social preference analysis, we keep in mind that our data on reporting, especially, are compatible with more psychological-style explanations, and we bring to bear evidence on the role of emotions in a final portion of our analysis.

The rest of the paper is organized as follows. Section 2 briefly summarizes related literature, and Section 3 describes our experimental design. Section 4 provides the theoretical predictions and hypotheses under both monetary payoff maximization and our illustrative social preference theory. Section 5 reports results. Section 6 concludes.

2. Literature

A large literature in economics discusses the social preferences that might lie behind certain deviations of behavior from those predicted in models of rational selfish agents with common knowledge of type. The mere belief that some agents may have altruistic, interdependent, or other "non-standard" preferences or emotions that interfere with selfishly rational decisions, or indeed even the belief that they believe that some others believe that some have such preferences (and so on), can change what is optimal for a strictly selfish agent. Kreps, Milgrom, Roberts and Wilson (1982) helped launch the formal study of such possibilities by assuming uncertainty regarding others' types and beliefs. The analysis by which we explain many decisions about cooperation in our experiment constitutes such a model, with agents who act rationally and self-interestedly yet depart from the predictions of conventional common

knowledge models due to beliefs that others' beliefs and actions may be non-standard. However, because we will also find decisions, especially decisions to engage in costly reporting, that are not compatible with material self-interest in view of the decision-makers' self-reported beliefs, the literatures on social preferences and emotions are also directly pertinent to our paper (for overviews, see Camerer and Loewenstein, 2003, Sobel, 2005).

The literature we see as closest to our topic is that on the experimental study of costly punishment, beginning with rejections in ultimatum game experiments (Camerer and Thaler, 1995; Camerer, 2003) and continuing into work on public goods games with punishment opportunities (Fehr and Gächter, 2000).[4] These latter experiments find that many subjects incur a cost to punish when there can be no material benefit to the punisher, and that the threat of punishment can reduce or eliminate incentives to free-ride.[5] Numerous papers extend these results (for reviews, see Gächter and Herrmann, 2009, Chaudhuri, 2011), including work focusing on the cost to the punisher (Anderson and Putterman, 2006, Carpenter, 2007), on punishment's effectiveness (that is, the punisher-punished cost ratio, see Nikiforakis and Normann, 2008), and on societal implications of costly punishment. Duersch and Servátka (2009) explore the effect of costly punishment in a prisoner's dilemma set-up, finding punishment to be less prevalent than in the literature on public goods games, but added punishment stages in the PD game otherwise appear absent from the literature.[6]

[4] Earlier, related studies, include Ostrom *et al.* (1992) and Yamagishi (1986).
[5] For example, Ertan *et al*. (2009), find that *ex post*, individual subjects earn more the more they contribute to the public good when opportunities to engage in costly punishment are available.
[6] A search for other experiments incorporating explicit punishment opportunities in prisoners' dilemma settings (leaving aside discussions of repeated game strategies) turned up contributions to a literature on the mathematics of evolutionary dynamics (for example, Dreber *et al*., 2008, and Rand *et al*., 2009), but none in the economics literature.

One way to understand costly punishment in a social dilemma game is to see punishers as having an aversion (beyond that associated with the pecuniary consequences alone) to others free-riding or defecting while they themselves cooperate. For such individuals, imposing an earnings reduction on free riders at a monetary cost to themselves delivers a utility gain that offsets their lowered money earnings. While such punishing can be thought of as resulting from a psychological trait of negative reciprocity (Hoffman, McCabe and Smith, 1998; Bowles and Gintis, 2004) perhaps linked to an emotional state of anger, it might also be rationalized or rendered mathematically tractable by a simpler framework of inequity aversion (Fehr and Schmidt, 1999).

Several scholars have researched costly reporting in the form of online product reviews (Dellarocas, 2003). Resnick and Zeckhauser (2002) look at the relationship between eBay reviews and sales, as well as the prevalence of and motivation behind reviewing. They find costly reporting to be frequent and suggest that the giving of feedback despite the absence of private material gain might be understood as the carrying out of a "quasi-civic duty" or as part of a "high courtesy equilibrium." Gregg and Scott (2006) find that eBay reviews are a major deterrent to fraud, helping to reduce asymmetry of information between buyers and sellers. Wang (2010) addresses the motivations behind leaving a review, specifically with respect to Yelp, a for-profit business review site. He finds strong evidence that social image and reviewer productivity are correlated.

We know of one paper, Gërxhani, Brandts and Schram (2013), in which costly reporting without clear private benefit plays an important part in an experiment. The authors study transmission of information about employee trustworthiness among employers, in one treatment making such transmissions anonymous so that direct reciprocity is ruled out as an incentive.

However, their players interact in the same condition for twenty periods, which can give rise to incentives for "reputation building."[7] Also, the reporting cost is quite small, and motivations to report, including their asymmetric applicability to reporting "bad actors," are not explored systematically or focused upon, as in our paper. Our paper is the first of which we are aware that studies costly reporting in a controlled laboratory setting where the reporting cost can be appreciable and private material gain is fully ruled out. Our simple design built on the canonical prisoners' dilemma game permits us to focus attention cleanly on the reporting decision.

3. Experimental Design

Our experiment consists of four main treatments with opportunities to report a counterpart's decision, in three of which reporting is costly. (An additional costly reporting treatment conducted by strategy method is discussed later.) In each treatment, subjects play two one-shot prisoners' dilemma games, each with different, anonymous, randomly selected participants. The payoff structures of the first and second games are identical and of equal money value to end-of-session earnings. The payoff table of each round in U.S. dollars is summarized in Fig. 1.[8]

A key feature of our design is that subjects decide in advance either to play the cooperation option (denoted X) or the defect option (denoted Y) in *both* games at the outset. Thus, in the instructions read by the participants, we write XX (YY) to represent the cooperate

[7] If players believe that beliefs that some people are "conditionally cooperative" are widely shared, then it can be privately beneficial to invest in promoting such beliefs by behaving cooperatively, since one may end up profiting from helping to build a "cooperative culture" within the group that may endure until "end game effects" set in. See, for example, the discussion in Palfrey and Prisbrey (1997).

[8] Note that payoffs were quoted in dollars and that no "lab currency" was used. It may be important to note that while predictions for the PD game are the same over a wide range of payoff configurations, the degree of "temptation" to defect, "fear" of being defected on, and the potential to gain from mutual cooperation relative to mutual defection, are impacted by the specific payoff structure, so that conclusions from an experiment with one payoff configuration may not extend, behaviorally, to alternative payoffs. See, for instance, Ahn *et. al* (2001).

(defect) option, duplicating the letter choice to indicate its play in two games. In our paper (but not subject instructions, which avoid such terms) we refer to a subject who chooses XX (YY) as a cooperator (defector), or occasionally XX- (YY-) chooser. Committing subjects at the outset to a single choice captures the notion that people have tendencies that they carry from interaction to interaction. Adoption of this feature greatly simplifies both analysis and reporting decisions, since it means that reporting, e.g., a defector, can be a reliable warning about the kind of agent the next partner will encounter. Of course, we need to take into consideration that imposition of this choice-for-both-games rule affects players' strategic calculations. Accordingly, this design feature of pre-commitment for two rounds of play ought not to be misconstrued as being a mechanism to force genuine type revelation in the sense of the theoretical literature. Indeed, we will show shortly that under some beliefs, it becomes rational for a strictly self-interested agent to select XX (cooperate).[9]

After being randomly matched with a counterpart, selecting between the two options, and being informed of the outcome of their first round interaction, each player in the reporting treatments decides whether to report the decision of her first counterpart. If a player is reported, then the second-period counterpart of the reported player is told what that player chose in the first round[10] and is given the option to change his initial choice of X or Y taking into account the report he received—this being the sole exception to the rule that an initial choice is binding for two rounds of play. Subjects know that they will certainly not play the game a second time with the same counterpart, so reporting in the hope that one might oneself be the beneficiary of the information is ruled out. The counterpart will also be told whether she was reported on by her

[9] Our analysis also implies, conversely, that there are conditions in which agents having genuinely conditionally cooperative or inequity averse preferences would choose YY (defect).

[10] That is, the computer delivers a truthful report. The potential issues of deciding whether to report truthfully and whether to believe a report that has been received are thus eliminated as concerns. We discuss the impact of this simplifying element in the conclusion.

initial partner, which determines whether the player whose initial choice was reported to her is in a position to select a new action.[11] For example, consider two players, A and B. Suppose that both A and B are paired with other randomly assigned partners (not each other) in the first game. Suppose also that A chooses XX (cooperate) and B chooses YY (defect). Now imagine that B's first round partner decides to report her (B's) decision to B's second interaction partner, namely A. Since B was reported, A has a chance to change his initial choice from X to Y so as to avoid being exploited by B. A is also free to stick to the choice of X. Suppose, finally, that A (the cooperator) is not reported by his initial partner. A is thus informed that B has no opportunity to change her choice, so A knows with certainty that B is playing Y in his interaction with her. In this example, A knows that switching to Y will protect him with certainty, that the choice would carry no danger of foregoing a mutual cooperation payoff, etc.

As another example, consider a cooperator (XX-chooser) C who learns that her second counterpart D had selected XX and has no opportunity to change his decision (D's initial counterpart did not report D's choice). C may wish to switch to Y to exploit D, but might decide to stick with X in order to avoid feeling guilty, experiencing disutility from advantageous inequality, etc. In the otherwise similar situation in which D can change his choice, C would have that information and would need to factor in her belief about the likelihood of D changing to Y (which may in turn be influenced by D's belief about the likelihood of C switching).[12] We

[11] To preserve maximum anonymity among the subjects in the experiment, those who had no opportunity to change their own choice were asked to answer a trivia question bearing no relation to the experiment, to keep number of computer clicks consistent across all those in the lab.

[12] While our design is one of finite repetition, we view the question addressed—that of lack of monetary incentive to engage in costly reporting—as relevant also to a world of indefinitely repeated interactions, since agents in such environments may also periodically need to seek new interaction partners and may avoid dealing again with an individual found to be opportunistic, but have no selfish material motive for incurring a cost to convey the information to others. Having only two interactions makes practical relatively large stakes in the lab for each interaction, while the fact that the report affects only one future interaction makes the motivational problem more challenging by limiting the punishment that reporting can inflict.

vary treatments by the cost of reporting—$1, $0.50, $0.05, or $0—referring to these as the High (Reporting) Cost (HC), Medium Cost (MC), Low Cost (LC) and No Cost (NC) treatments, respectively.

Subjects are also asked, after their own choice of XX or YY, for their beliefs about the percentage of their peers choosing XX, and in the four main treatments they are asked—after their reporting decisions—for their beliefs about the percentage of defectors and that of cooperators who will be reported. So as not to raise its salience too much, we do not tell subjects about the presence of the belief elicitation tasks before they make the corresponding choices. Subject are asked for their expectations regarding behaviors of *other* participants only (themselves not included), to avoid hedging. Eliciting beliefs is incentivized by offering a $1 bonus payment for guesses that are within 5 percentage points of the actual percentage. Including belief elicitation in our design permits us to explore the driving forces behind the subjects' decisions. At the end of their session, subjects are also asked about emotions potentially affecting their reporting decisions: (i) their level of anger toward their initial partners and (ii) their feelings of obligation to help the third party in the second round via reporting.

4. Theoretical Predictions without and with Social Preferences

Although our main focus is on costly reporting, which we recognize from the outset to be ruled out in our setting for traditional economic actors with common knowledge of type, we develop here predictions of subject decisions with regard to both the report/don't report and the cooperate/defect choice. We begin our analysis with the extreme assumption of strictly selfish preferences, rationality, and common knowledge, then relax the common knowledge assumption to allow for beliefs that others may report and cooperate. Finally, we relax the selfish preference

assumption to allow that some may have social preferences capable of explaining actual costly reporting and some decisions to cooperate contrary to material self-interest. We leave consideration of emotional factors to be discussed when we view the experimental results.

4.1 Common knowledge of rationality and self-interest

The standard theory predictions in the experiment (assuming rationality, self-interest, and common knowledge) are straightforward. In the HC, MC and LC treatments, it is never payoff-enhancing to report, since reporting is costly and players are never matched with the same partner twice. With the probability of being reported on being zero, the prediction for self-interested players with common knowledge is the same in each of our one-shot games as in any single-play prisoner's dilemma, i.e. a subject will always choose to defect. Being forced to select a decision for both games simultaneously thus makes no difference. In the NC treatment, subjects should be indifferent about reporting their counterparts' decisions, since reporting has no effect on own payoff. While this might be supposed to generate a 50/50 chance of a given subject's action being reported and of her second counterpart accordingly having a chance to change his decision on the basis of that information, defecting at the outset (and, if offered a choice, in the second game as well) is still the dominant strategy unless a substantial proportion of others are expected to cooperate and reporting decisions are non-random (see the next sub-section).[13] We conclude, then, that under the standard assumptions of payoff-maximization, rationality, and common knowledge, there is no reporting at positive cost, all games played end up with mutual defection, and subjects obtain the (Y,Y) payoff of $5 in each of their two games.

[13] Assuming, alternatively, that subjects always report when it is cost-free to do so, would leave YY the predicted first choice and Y the free second choice of those receiving reports, thus having no impact on how the PD games themselves are played.

***H**ypotheses with **S**elf-Interest, **R**ationality and **C**ommon Knowledge (H-SRC):*

In the HC, MC and LC treatments, each subject chooses not to pay to report choices of her first interaction partner, while reporting occurs randomly in the NC treatment. Subjects in all treatments choose YY (defect). A subject having the opportunity to make a free second choice (which, by the above, occurs only in the NC treatment) always selects Y.

4.2 Dropping common knowledge

Many laboratory decision experiments have found that some people behave as if having other-regarding preferences such as inequity-aversion, altruism or reciprocity. Even if all subjects are strictly self-interested and rational, the belief that others might behave pro-socially and/or that others believe such types exist, can move behavior dramatically. In this sub-section, we consider what *beliefs* can make cooperation selfishly rational. We leave the effect of own social preferences on individuals' actions to be considered in sub-sections 4.3 and 4.4. Since subject *i* remains a self-interested payoff maximizer in the present sub-section, the prediction that *i* will never pay to report others remains unchanged from *H-SRC*.

Let a_i be the fraction of subjects that *i* believes will cooperate (select XX), b_i the fraction she believes will report a cooperating counterpart, and c_i the fraction she believes will report a defecting one ($0 \leq a_i, b_i, c_i \leq 1$). We solve for the conditions under which *i* selects XX or YY under a range of assumptions *i* might make about the 2nd game behavior of a participant who receives a report.

Consider two extreme assumptions regarding beliefs about any free 2nd game choices that become available. We label these "pessimistic" and "optimistic," respectively. Under the pessimistic assumption, decision-maker *i* assumes that an individual free to revise her second choice always selects Y in line with selfish preferences. Under the optimistic assumption, *i*

assumes that cooperatively-oriented participants (XX-choosers) will stick with X given an opportunity to make a fresh choice provided that they are informed that they are meeting another subject who chose XX. We could distinguish predictions based on whether the XX-chooser being met with can also change her choice, and we could vary the optimism of the belief, but to provide a simple alternative to the pessimistic assumption, we define the optimistic one as assuming that XX-choosers stick with their choice of X when knowing they have met one another, period. Even this belief is bounded in optimism in that YY-choosers free to revise their second choice are still assumed always sticks with Y.[14] Making some assumption about players' beliefs concerning what actors do if facing a free 2nd choice is necessary to our analysis, and we think it reasonable to suppose that actual beliefs tend to fall somewhere between the two just laid out.

As shown in Appendix A.1, under the pessimistic 2nd game assumption, we obtain:

$$\text{selfish player } i \text{ cooperates (defects) if } 5a_i(c_i - b_i) + c_i > 2\ (< 2) \qquad (1)$$

By contrast, under the optimistic 2nd game assumption, as shown in Appendix A.2, we obtain:

$$\text{selfish player } i \text{ cooperates (defects) if } 5a_ic_i + a_ib_i + c_i > 2\ (< 2) \qquad (1')$$

By inspection, under the pessimistic assumption, when defectors (YY-choosers) are believed to be reported more often than are cooperators ($c_i > b_i$), both a higher expected share of cooperators (a_i) and a higher expected share of defectors being reported (c_i) increase the likelihood of cooperation being the payoff-maximizing choice, while a greater expected share of cooperators being reported (b_i) lowers it. Intuitively, higher a_i reduces the net expected gain from defecting when defectors are differentially reported ($c_i > b_i$), higher c_i raises the likelihood of a mutual

[14] Our data show some (but not all) cooperators stick to cooperation when having the report of meeting a cooperator, but no defector switches to cooperation in this situation, so greater optimism than our optimistic assumption does not strike us as worthy of much attention.

defection outcome in the second interaction, and higher b_i reduces the likelihood of mutual cooperation in that interaction (by the assumption that any subject receiving a report will defect).[15] But under the optimistic assumption, rather than exert negative weight as in criterion (1), a higher share of cooperators being reported—b_i—if anything encourages cooperation, as selfish player *i* can exploit such a cooperator. Note that a higher fraction of subjects chooses to cooperate based on criterion (1') compared with criterion (1). If some subjects choose to cooperate even though criterion (1') does not hold, their choices either result from error or from non-selfish motives, as explored shortly.

***H**ypotheses assuming Self-interested Choice allowing for Belief in **S**ocial **P**references among **O**thers (H-SPO):*

Suppose that subject i is a material payoff maximizer, but believes that others might cooperate and pay to report (cooperators, defectors) with probabilities a_i, b_i, and c_i possibly > 0. Then i will never pay to report her first interaction partner, and will randomly report or not report if reporting is cost free. Subject i will choose cooperation (defection) if $5a_i(c_i - b_i) + c_i > 2$ (< 2), assuming that i has "pessimistic" beliefs about cooperators' free 2nd choices, and will choose cooperation (defection) if $5a_ic_i + a_ib_i + c_i > 2$ (< 2), assuming that i has "optimistic" beliefs about those choices. Subject i will always choose Y if able to make a free 2nd choice.[16]

4.3 Social preferences and decisions to report

[15] Since the LHS of equation (1) must exceed 2 for cooperating to be selfishly rational, it is in fact necessary that the amount by which c_i exceeds b_i and the degree to which a_i exceeds 0 must both be non-negligible if choice of XX is to be payoff-maximizing.

[16] Note that by writing conditions (1) and (1') with strict inequalities, we assume that subjects whose beliefs render them rationally indifferent between XX and YY will select the latter, if strictly self-interested. This means that when analyzing our experiment results in section 5, we place the few cases in which (1) or (1') hold with exact equality in the set of observations for which an observed choice of XX requires a social preference explanation.

If our focus were on cooperation choices, we could perhaps end our analysis with prediction H-SPO, since if all involved were indeed selfish and rational, then most decisions to cooperate could perhaps be attributed to incomplete knowledge of others' types and beliefs (e.g., mere beliefs of $c_i > 0$ could theoretically drive cooperation though none are actually motivated to pay to report). But the main focus of our paper is costly reporting, and we conjecture that such reporting will indeed occur. Therefore, something, perhaps a social preference or emotion, is needed to explain that behavior. Suppose that the explanation for reporting is a social preference; in particular, suppose that the potential reporter *i* has the inequity-averse preferences proposed by Fehr and Schmidt (1999):

$$u_i(\pi_i | \pi_j) = \pi_i - \alpha_i \cdot max\{\pi_j - \pi_i, 0\} - \beta_i \cdot max\{\pi_i - \pi_j, 0\}, \quad (2)$$

where $\alpha_i \geq \beta_i \geq 0$, meaning that aversion to inequality that is unfavorable to the decision-maker (reflected in weight α_i) is at least as strong as that to favorable inequality (reflected in weight β_i), and that the decision-maker never values the latter ("aheadness") for its own sake (β_i takes no negative values). There are four possible situations under which reporting might occur:

Case 1: subject *i* cooperates and learns that her counterpart has also cooperated.

Case 2: subject *i* cooperates and learns that her counterpart has defected.

Case 3: subject *i* defects and learns that her counterpart has cooperated.

Case 4: subject *i* defects and learns that her counterpart has also defected.

Assume, further, that the only other individual, *j*, whose payoff π_j affects u_i if α_i (and perhaps β_i) > 0, is the first interaction partner of decision-maker *i*, with respect to whom *i*'s decision to

engage in reporting is made.[17] Then it can be shown (details are in Appendices A.4) that costly reporting will occur

$$\text{in Case 1 if } 6a_i > \rho,\ (\alpha + \beta)b_i - \beta > 0 \text{ and } (6a_i - \rho) > \rho/[(\alpha + \beta)b_i - \beta] \quad (3a)$$

$$\text{in Case 2 if } (6a_i - \rho) > \rho/\alpha \quad (3b)$$

$$\text{in Case 4 if } 6a_i > \rho,\ (\alpha + \beta)c_i - \beta > 0 \text{ and } (6a_i - \rho) > \rho/[(\alpha + \beta)c_i - \beta] \quad (3c)$$

where ρ is the reporting cost.[18] The analysis in the Appendix shows that the conditions for Case 2 and Case 4 hold regardless of whether *i* applies the pessimistic or the optimistic assumption about free 2nd choices, while the condition for Case 1 applies only when *i* makes the pessimistic assumption; if she makes the optimistic assumption instead, *i* will never pay to report. As for Case 3, the analysis indicates that a defector *i* will never report a cooperating counterpart if $\beta_i < 1$, a restriction that Fehr and Schmidt find applicable when laying out the sets of values that can describe populations, based on their analysis of several kinds of bargaining experiment data.[19] Since abnormally high β values are if anything less likely to be found among defectors (see section 4.4, below), we conclude that costly reporting of cooperators by defectors will rarely if ever occur.

Condition (3b) indicates that the more averse to disadvantageous inequality is the cooperator (the higher her α_i), the more others she believes to have chosen to cooperate (the higher her a_i), and the lower is the reporting cost ρ, the more likely she is to report a defector.

[17] Alternative assumptions, for instance that *i* cares about inequalities with respect to all other participants, cannot be ruled out, but where *i* stands relative to the individual with whom she has just played and about whom she makes the reporting decision seems likely to be most salient. *i*'s potential concern for *j*'s next partner, whom *i* may wish to warn or at least inform of *j*'s type, is also a plausible concern and amenable to analysis using the Fehr-Schmidt model, but we focus on *i*'s concern with *j*, partly for reasons discussed in footnote 40.

[18] *i* is indifferent between reporting and not reporting if the right hand inequality in each line holds instead with the equals operator.

[19] Specifically, Fehr and Schmidt conclude that about 30% of individuals have $\alpha = \beta = 0$, about 30% have $\alpha = 0.5$, $\beta = 0.25$, 30% have $\alpha = 1$, $\beta = 0.6$, and 10% have $\alpha = 4$, $\beta = 0.6$. See also Table 1 in Fehr and Schmidt (2010).

Aversion to advantageous inequality, β, plays no role in this decision. In addition, the condition implies that all inequity-averse cooperators report defectors at zero cost (in the NC treatment), as long as $a_i > 0$.

Although conditions (3a) and (3c) appear slightly more complex, it is easy to show that if distributions of types (α and β values) are no different in each case, then the threshold belief a_i required for costly reporting is lowest for Case 2.[20] In conditions (3a) and (3c), we also have the interesting further implication that costly reporting is conditional on the belief that others do it (a cooperator [defector] is more likely to report another cooperator [defector] if she has a high belief b_i [c_i]). As for relative frequency of reporting in cases 1 and 4, the two conditions show reporting to be less likely in Case 4 than Case 1 if $c_i > b_i$, an intuitively appealing idea that turns out to be strongly supported by our experimental data.[21] Adding to this the finding that reporting is not predicted by any individual making the optimistic assumption about free 2nd choices, we arrive at the prediction that, assuming sufficient variation of belief a_i not systematically linked to preference type, costly reporting should be most common in the case of cooperators meeting defectors (Case 2), with the cases of defectors meeting defectors (Case 4) and cooperators meeting cooperators (Case 1) following in that order.

[20] The right hand inequalities of (3a), (3b) and (3c) are identical apart from denominators which have the forms $[(\alpha + \beta)b_i - \beta]$, α, and $[(\alpha + \beta)c_i - \beta]$, respectively. It is easy to show that the denominator is smaller, thus creating a higher hurdle that the left hand side of the inequality must exceed, in (3a) and (3c) than in (3b). As for the assumption that the type distribution will not vary among cases, it is clear that the same types must end up in cases 1 and 2, on average. Those in Case 4 may tend to have somewhat lower values of β_i on average (see section 4.4), but that if anything strengthens the case for expecting the denominator on the right hand side of (3c) to be larger than its counterpart in (3a) (see footnote 21).

[21] That reporting will be more frequent in Case 4 if $c_i > b_i$ is shown by comparing the denominators of (3a) and (3c) and seeing that their relative sizes to hinge on the sizes of beliefs b_i vs. c_i. Inspecting the beliefs reported by our subjects shows that less than 32% of subjects believed that the fraction of cooperators reported would exceed the fraction of defectors reported.

Hypotheses on Reporting Decisions with Inequity-Averse Social Preferences (H-R-SP):

(i) Costly reporting due to inequity averse preferences is most likely in the case of a cooperator meeting a defector, followed by the case of a defector meeting a defector, with reporting of cooperators by cooperators still less common and that of cooperators by defectors rarely if ever occurring; (ii) defectors (cooperators) are more likely to report their defector (cooperator) counterpart the greater the share of others they believe report (the greater is their belief c_i [b_i])—i.e., individuals tend to report when they believe others tend to report; (iii) the higher is belief a_i, and the lower is reporting cost ρ, the larger the share of individuals who will engage in costly reporting.

4.4 Social preferences, beliefs, and decisions to cooperate

The effects of inequity aversion or other social preferences can lead not only to costly reporting, but also to decisions to cooperate despite failure of the relevant inequality (1) or (1') to hold. Combining beliefs a_i, b_i and c_i, the pessimistic assumption about free 2nd choices that underlies inequality (1), and inequity-averse utility function (2), we can derive (see Appendix A.3) optimizing condition:

$$\text{cooperate (defect) if } 5a_i\cdot(c_i - b_i) + c_i > (<)\ 2 - (8 - 5a_i\cdot(c_i - b_i) - c_i)\cdot(a_i\cdot\beta_i - (1 - a_i)\cdot\alpha_i) \qquad (4)$$

If instead, the more optimistic assumption about free 2nd choices underlying (1') is used, the optimizing condition becomes:

$$\text{cooperate (defect) if } 5a_i\cdot c_i + a_ib_i + c_i > (<)\ 2 - (8 - 5a_i\cdot c_i - a_ib_i - c_i)\cdot(a_i\beta_i - (1 - a_i)\alpha_i) \qquad (4')$$

We see that the LHS of (1) [of (1')] need not reach the threshold of 2 if the decision-maker has great enough aversion to inequalities that favor her, is not too much more averse to

disadvantageous than to advantageous inequalities, and expects a large proportion of others to cooperate (has large a_i).

***H**ypothesis on **C**ooperating with Inequity-Averse **S**ocial **P**references (H-C-SP):*

A subject i who has inequity-averse preferences will be more likely to choose XX the more others she expects to cooperate (the higher her belief a_i), the more averse to advantageous inequality she is (higher β_i), and the less averse to disadvantageous inequality she is (lower α_i).[22]

4.5 Anticipating others' cooperation

A potential problem that all of conditions (1) to (4') share is that they can show cooperating or reporting to be individually optimal only for individuals who for some reason believe that some others will cooperate ($a_i > 0$). For example, the minimum level of a_i that is required in order to make cooperating selfishly rational, according to (1), occurs when $c_i = 1$ and $b_i = 0$, in which case a_i must be at least 0.2. (3) indicates that a_i must be greater than $((1/\alpha_i) + 1)\cdot\rho/6$ for it to be rational for inequity-averse agent *i* to report her defecting counterpart. These conditions may thus arguably suffer from circularity: if enough individuals believe that enough individuals believe that enough others will cooperate (a_i is large enough) to make cooperating individually optimal, the beliefs in question can become self-fulfulling, but how the beliefs get started is unclear. Possibly the initial belief can rest on the idea that most people are socialized to adhere to a maxim such as the Golden Rule and that because that rule implicitly calls for cooperating, many will cooperate, or at least enough others will believe that enough others will

[22] With no possibility of reporting and of changing 2nd play decisions, inequity averse individuals having high a_i, β_i, and β_i/α_i can prefer to cooperate, because they have a higher expected subjective cost from gaining at the advantage of a cooperative counterpart than of losing at the hands of an opportunistic one, assuming that the likelihood of meeting the latter is not high.

cooperate so as to achieve the relevant hurdle for *a* by the decision-makers in question.[23] Efficiency preferences, for example positive valuation of aggregate earnings as in the model of Charness and Rabin (2002), could also lead some to cooperate. Another line of argument is that the belief that others believe that some minimum number of others will cooperate rests on no particular conjecture about those others' preferences, but is simply a "rational expectation" of how many people happen to cooperate, on average. Without knowing exactly *why* people cooperate, one can see in existing experimental one-shot PD games that initial cooperation rates of around 30 – 40% are common.[24]

5. Results and Analysis

5.1 Overview of the experiment

A total of 172 students (152 in the four main reporting treatments, 20 in the strategy method treatment) participated in the experiment sessions in 2013 and 2014 at the University of Michigan in Ann Arbor.[25] 58.7% of subjects (101) were female. No subject participated in more than one session, and the sessions lasted about an hour on average. The experiment was programmed in z-tree (Fischbacher, 2007). All instructions (for the example of the LC treatment, see the Appendix) were neutrally framed, avoiding terms such as "cooperate," "trust," etc. Subjects had to answer a number of control questions to confirm their understanding of the experiment. Communication between subjects was not permitted. Average earnings were $20.84, including a $5 participation fee, with a standard deviation of $4.16.

[23] Dal Bó and Dal Bó (2014) show that exposing subjects to a statement of the Golden Rule increases their cooperation rates in a public goods dilemma.
[24] See for example, Cooper *et al.* (1996), Ahn *et al*. (2009), and Dal Bó *et al*. (2010).
[25] Sessions were conducted by Kamei while he was Assistant Professor at Bowling Green State University. All subjects were recruited from the University of Michigan experimental lab's subject pool using solicitation messages via ORSEE (Online Recruitment System for Economic Experiments). An additional treatment without reporting opportunities is not reported, to conserve space.

We begin with the focal choice: costly reporting. The upper middle panel of Fig. 2 shows that some 45 to 65% of cooperators who encountered a defector chose to report when costly. There is little sign of correlation between cost and reporting incidence, but reporting occurs considerably more often (almost 90% report) when the cost is zero (NC treatment) than when it is positive (an overall average of 58.6% report in LC, MC and HC). There is also some, but considerably less, costly reporting of cooperators by cooperators and of defectors by defectors, and no costly reporting of cooperators by defectors. Overall, an average of 8.0% of subjects in the cooperator-cooperator, defector-cooperator and defector-defector situations choose to report.

Result 1: (a) Costly reporting of defectors by cooperators is common (almost 59% report), inconsistent with H-SRC and H-SPO but consistent with H-R-SP. (b) Costly reporting in other cases is significantly less common (overall, 8%), consistent with H-R-SP. (c) Somewhat contrary to H-R-SP, the frequency of costly reporting does not vary systematically with its cost, except insofar as (d) there is significantly less reporting at a positive than at a zero cost, consistent with H-R-SP.

Turning to cooperation decisions and expectations, the left panel of Figure 2 shows that regardless of the presence of reporting costs and their size if present, around 50 to 60% of the subjects choose XX in each of the main treatments.[26] The diamonds and triangles in the same panel indicate that cooperators' average expectation regarding the fraction of others who would choose cooperation was significantly higher (around 70%) than that of defectors (around 40%). The difference is significant with $p < 0.001$ (see Appendix Table B.3). With respect to reporting, the upper right panel shows that the average cooperator believed more defectors would be

[26] The fractions of cooperators are not significantly different between any two treatments, according to two-sided, two-sample tests of proportions. See Appendix Table B.1 panel (C).

reported than did the average defector, a difference significant for LC subjects and in all costly reporting treatments pooled, with $p < 0.01$ (see Appendix Table B.4).

The displayed frequencies of cooperation and reporting are clearly inconsistent with H-SRC, which, based on the assumption of rational selfish individuals with common knowledge, predicts neither costly reporting nor cooperation.[27] Before returning to the reporting decisions, which are our central focus but which require explanation by social preference or behavioral factors, we explore first the extent to which cooperation choices can be explained by own self-interest plus beliefs that others have social preferences (H-SPO).

5.2 Do beliefs support rational cooperation?

Table 1 shows overall average beliefs, average beliefs disaggregated to distinguish cooperators and defectors, and the actual *ex post* shares corresponding to each of beliefs a_i (proportion cooperating), b_i (proportion of cooperators reported) and c_i (proportion of defectors reported) by treatment. It also shows the average value of the term $5a_i(c_i - b_i) + c_i$, which is the LHS of inequality (1), and the average value of the term $5a_ic_i + a_ib_i + c_i$, which is the LHS of inequality (1'). If an individual's expectations cause the relevant expression to exceed 2, it would raise her payoff to select XX, under the "pessimistic" ((1)) and "optimistic" ((1')) assumptions about free 2nd round choices, respectively. We see that the overall average value of the first expression falls substantially below the cooperation threshold level of 2, but that of the latter is above that threshold in all treatments, and well above it in most. The data also indicate that the within-treatment average rises monotonically as reporting cost falls, mainly because

[27] We performed binomial probability tests for the conservative null hypothesis that the probability of choosing XX equals 5%, assuming that errors occur with a probability of 5%. This hypothesis was rejected in each treatment. We also performed binomial probability tests for the conservative null hypothesis that the probability that cooperators report the initial choices of their YY-choosing partners equals 5%, assuming that errors occur with a probability of 5%. This hypothesis was also rejected in each treatment. See Appendix Table B.2.

while actual reporting did not monotonically increase as its cost fell, the expectation of reporting (*c*) did (see rows *c* of both the upper ALL SUBJECTS portion and the middle COOPERATORS portion of the table).

Although cooperating is never selfishly rational for a subject holding the average belief and pessimistic 2nd game assumption, this is not true of many individual cooperators, who had substantially more optimistic beliefs a_i and c_i than did the average defector. In the row labeled "LHS of (1)" in the COOPERATOR portion of Table 1, we find that the expression's value did on average exceed 2 for cooperators in two treatments, LC and NC. The corresponding row for defectors shows the expression's value to be well below even 1 in all four reporting treatments, indicating that the average defector's choice of YY was well in line with the prediction for selfishly rational agents.

We further disaggregate the data to the individual level in order to see exactly what proportion of cooperators and defectors' choices can be rationalized by payoff-maximization conditioned on own beliefs. Appendix Table B.3 details that 26.1%, 31.6%, 61.9% and 43.5% of cooperators had beliefs making choice of XX payoff-maximizing under the pessimistic 2nd game assumption in the HC, MC, LC and NC treatments, respectively. If the optimistic 2nd game assumption is applied, 43.5%, 63.2%, 90.5% and 78.3% of cooperators could have been choosing XX to maximize own payoff in the HC, MC, LC and NC treatments, respectively. Assuming that many of those who chose cooperation made intermediate assumptions about free 2nd game choices, i.e. assumptions lying somewhere between the "pessimistic" and "optimistic" ones, frequency of cooperation out of self-interest would lie between those numbers. Among the defectors, 100%, 89.5%, 94.7%, and 92.3% had beliefs making their choice of YY payoff-

maximizing, with the pessimistic assumption.[28] Importantly, we see that while a substantial majority of all choices between cooperation and defection (63.8% with the pessimistic 2nd game assumption, 79.6% with the optimistic 2nd game assumption) can be explained by self-interested rationality, that approach explains a considerably smaller share of cooperators' than of defectors' decisions. Put differently, the overwhelming majority of subjects who departed from the self-interest assumption of H-SPO displayed a bias towards cooperation. This suggests that the assumption of selfishness may need to be relaxed at least for some cooperators, as we further explore below.

Result 2: (a) Almost two-thirds (four-fifths) of choices between cooperation and defection are consistent with own payoff maximization using the pessimistic (optimistic) 2nd game assumption, consistent with H-SPO; (b) Almost all choices that deviate from self-interest under these alternative assumptions are decisions to cooperate.

5.3 Predicting reporting decisions with inequity-averse preferences

Whereas the greater part of the cooperation that we find in our data is potentially explicable as self-interested responses to beliefs about others' cooperation and reporting, our design intentionally rules out any such motivation for costly reporting, which is our primary focus. As we saw in section 5.1, substantial numbers actually report defectors at a cost to themselves despite complete absence of potential material benefit, paralleling the way in which substantial numbers pay to punish free riders in one-shot voluntary contribution dilemmas.

[28] An additional question that can be asked is whether cooperation would have been profitable for a subject having *ex post* accurate beliefs about the shares of cooperators and defectors who would be reported by their partners. The answer could help to assess whether cooperation could be sustainable in a setting like ours having additional rounds of play and accompanying opportunities to learn about reporting frequency. We provide the relevant analysis in the bottom portion of Table 1. We show there that in none of the three costly reporting treatments does the LHS of (1) or (1') based on *ex post* accurate beliefs reach the necessary threshold of 2. However, the value is rather close to the threshold in the HC treatment.

Specifically, in the costly reporting treatments as a group, some 45% to 65% of cooperators incurred costs of between \$0.05 and \$1.00—the latter constituting about 7% of the earnings predicted by traditional theory, or 10% of those earnings net of the show-up fee—to report a defecting counterpart. Depending on treatment, up to 20% of defectors meeting defectors, and likewise up to 20% of cooperators meeting cooperators, engage in costly reporting, but in most treatments the share reporting is closer to 10%, and that share is always substantially, and often statistically significantly, below the share of cooperators reporting defectors that choose to report (see again Appendix Table B.1). Overall, in the three treatments in which reporting is costly, 58.6% of cooperators meeting defectors, 16.6% of defectors meeting defectors, 8.8% of cooperators meeting cooperators, and 0% of defectors meeting cooperators, pay the cost to report their counterpart, an ordering of frequencies exactly matching part (i) of H-R-SP.[29]

Although we can't identify individual utility function parameters in order to predict which subjects will report their counterparts, we can estimate the proportion who would be expected to report based on conditions (3a), (3b), (3c) and self-reported beliefs a_i, b_i and c_i, if we assume that each subject has the same likelihood of belonging to each of the four preference types identified and assigned estimated population proportions by Fehr and Schmidt, in precisely those proportions.[30] The calculations thus made imply that about 68.9% of the cooperators who encountered defectors, 13.5% of the cooperators who encountered cooperators, and 14.2% of the defectors who encountered defectors would engage in reporting at the costs obtaining in their treatments given their beliefs and the prevalence of each type. These predicted shares are rather

[29] The percentage of defectors being reported by cooperators is significantly larger than the percentage being reported in any other pairing of actions (defectors being reported by defectors, cooperators being reported by cooperators, and cooperators being reported by defectors) according to two-sample z-tests of proportions using data of the three costly reporting treatments.

[30] Each identifiable subject in each meeting case, that is, has a 30% chance of having $\alpha = \beta = 0$, a 30% chance of having $\alpha = 0.5$, $\beta = 0.25$, etc. See again note 19, above.

similar to the shares actually reporting (see above), the similar differences between the high share for Case 2 reporting and the low shares for Case 1 and Case 4, in estimate and reality, being especially remarkable.

We can also estimate multivariate regressions to check for patterns consistent with reporting criteria (3a) – (3c). In Table B.8 of the Appendix, we show estimates of simple linear regressions in which the independent variables are the values of the three beliefs and the two treatment dummies, which control for reporting cost. The regressions for Case 1 and Case 4 show partial consistency with conditions (3a) and (3c) in that the belief variables b_i and c_i obtain positive and significant coefficients in their respective estimates. These coefficients suggest a sort of "conditional cooperativeness" with respect to reporting: the higher the fraction of others a subject believes report a player who behaves like her counterpart (one who cooperates, in Case 1, one who defects, in Case 4), the more likely that the subject herself pays to report. The estimate for Case 2 suggests a similar sort of conditionality: among cooperators who meet a defector, those believing that a higher share of defectors are reported are more likely to report, themselves.[31] The coefficients on the expected share cooperating (a_i) and on treatment dummy variables are insignificant (in one case marginally significant), however, failing to support expectations based on conditions (3a) – (3c) that frequency of reporting would be increasing in a_i and decreasing in ρ.[32] H-R-SP is supported, with respect to the relationship between cost and reporting frequency, only insofar as there is far more reporting at a cost of zero than at a positive cost, in general, and at any of the individual costs, in particular (see Panels (A) and (B) of

[31] There have been numerous behavioral findings of tendency to perform a pro-social or cooperative act conditional on beliefs that others do so; see Fischbacher and Gächter (2010) for conditional contributing in public goods games and Kamei (2014) for conditional costly punishing in public goods games with punishment opportunities.
[32] Failure of belief a_i to obtain a significant positive coefficient, whereas c_i obtains one, in the regression for the Case 2 data, is inconsistent with condition (3b), which implies that a_i rather than c_i should be significant. The estimated coefficient on a_i is insignificant, however, even in specifications that exclude b_i and c_i terms.

Appendix Table B.1).[33] Subjects' self-reported expectations are also ones of greater reporting in treatments with lower reporting cost, although that ordering does not emerge in practice.

Result 3: (i) The percentages of subjects engaging in costly reporting are well predicted using Fehr and Schmidt's estimates of the prevalence and strength of aversion to disadvantageous inequality and our subjects' self-reported beliefs about frequency of cooperation. Specifically, costly reporting of defectors by cooperators is by far the most frequent case, with costly reporting of cooperators by defectors (predicted to occur rarely if ever) not observed and reporting in the remaining two cases relatively infrequent. (ii) Sensitivity of reporting to its cost is limited, in our data, to the presence of significantly more reporting at cost 0 than at costs \$0.05, \$0.50 and \$1.00. [34]

5.4 Cooperating out of a social preference

Even with the optimistic beliefs used to derive condition (1'), the cooperate/defect choice of between 9.5% of cooperators in the LC treatment and 56.5% of cooperators in the HC treatment is inconsistent with strict self-interest, and thus suggests either error, limited calculating ability, or preferences not based on own money payoff alone. A social preference explanation is consistent with most of these decisions.

Consider, again, a subject having the inequity-averse preference defined by Eq. (2). High enough values of the advantageous inequity aversion parameter β relative to disadvantageous

[33] The fact that reporting is far greater at zero cost than at a money cost as low as \$0.05 might reflect a psychological distinction between money and time costs, or a peculiarity of the zero cost as discussed, for example, by Shampanier *et al*. (2007). This suggests that there could be a substantial numbers of individuals willing to provide online reviews with what they treat psychologically as spare time, but who would be deterred if even a small monetary cost were involved, but we are unaware of any tests of this conjecture.

[34] It should be noted that with the limited variation in values of α_i and β_i assumed by Fehr and Schmidt (see note 19 above), and with the high expectations a_i of most cooperators, little sensitivity to reporting cost within the range including in our data is predicted. For example, our calculations predict that 70% of cooperators will report a defector in both treatments LC and MC, with a drop only to 66.7% predicted to report defectors in HC.

inequity aversion parameter α can drive choice of XX provided that the proportion of others believed to be choosing XX is sufficiently high. The most conducive configuration of values discussed by Fehr and Schmidt is $\alpha = 1$, $\beta = 0.6$, values they estimate to characterize 30% of randomly chosen individuals and ones at which an individual would trade a $1 reduction in disadvantageous inequality for $1 of own income while valuing a $1 reduction in inequality that advantages her at 60 cents of own income. As reported in our appendix Table B.3, roughly 12% of cooperators in the main treatments for whom neither (1) nor (1') exceed the required threshold value of 2, would have found choosing XX to be utility-maximizing given those parameter values and given their beliefs a_i, b_i and c_i. Inequity aversion can explain most of remaining decisions to cooperate, but only with higher β values than believed likely by Fehr and Schmidt.[35]

Result 4: Most cooperation choices that are not consistent with selfish rationality can be explained by aversion to advantageous inequality, consistent with H-C-SP. 38% of the cases explicable by inequity aversion are fully consistent with Fehr and Schmidt's estimates of type distribution, while the remainder require degrees of aversion to advantageous relative to disadvantageous inequality in excess of the levels specified in Fehr and Schmidt's typology.

[35] If $\alpha = \beta = 1$, for instance, four more subjects (around 5% of the cooperators in the main treatments, overall), can be added to the set of subjects for whom that decision is utility-maximizing. A still more extreme assumption of $\alpha = 0$, $\beta = 1$ (the individual would pay nothing to eliminate disadvantageous inequalities but would pay $1 to reduce by $1 the gap between a less well-off individual and herself) could explain an additional 14% of choices to cooperate. Overall, then, while 39.7% of cooperation choices are explicable by self-interest with the pessimistic 2nd choice assumption and another 25.4% are explicable by self-interest with the optimistic 2nd choice assumption, an additional 14% of cooperation is explicable by inequity-averse preferences and the parameters of Fehr and Schmidt (1999), and almost all of the remaining cases (about 19% of the total) are explicable by inequity aversion more strongly biased against advantageous inequality.

5.5 Role of emotions

Although the psychological underpinnings of reporting and the possible role of emotions were mentioned in earlier sections, our analysis in Section 4 and thus far in the present section has focused on the application of formal utility models to explaining subject behaviors. In this sub-section, we consider evidence that emotions may have played a part in our subjects' decisions.[36]

One source of evidence to this effect is an additional treatment we conducted that resembles the MC treatment in all respects except for using what experimental economists call the "strategy method"—that is, rather than having subjects decide whether to report their initial counterpart's choice after learning what that choice was, subjects are asked to decide in advance whether they wish to report their counterpart if the participant with whom they are matched for their first interaction turns out to have selected XX, and likewise whether they wish to report their counterpart if he or she turns out to have selected YY.[37] Of the 20 subjects that participated in this strategy method session, 11 selected XX, similar to the 50% share making that choice in the MC treatment. Thanks to the strategy method set-up, we got decisions from all 11 about whether they would report if meeting a YY-chooser.[38] Only 2 of the 11 (i.e., 18%) chose to pay the required $0.50 to report their first counterpart if that person chose YY. In the MC treatment, with the same payoffs and cost but using the sequential method, there were 9 who chose XX and met a counterpart that chose YY. Of those 9, 6 chose to engage in costly reporting upon learning

[36] Social preference models, despite their rational choice formalism, are not necessarily incompatible with more emotion-referencing neuro-biological explanations of behavior; see Fehr and Camerer (2007).

[37] These reporting decisions were taken after each subject had made her own choice between XX and YY and had indicated what percentage of others she expected would choose XX. Following the reporting decisions, subjects stated their beliefs about others' reporting decisions under each contingency. To further minimize the impact of the expectation elicitation process, subjects were not informed in advance of the fact that expectations were to be elicited.

[38] Thus, a single strategy method session generated the same number of decisions by XX-choosers regarding whether to report a YY-chooser as did the two MC treatment sessions.

of their counterpart's action, i.e. 66.7%. A two-sided two-sample z-test of proportions says that the 18% and 66.7% proportions are statistically significantly different from each other with p = 0.028. This difference suggests that "hot" emotion (Loewenstein, 2000) may have played a role in reporting in our main treatments.

Another piece of evidence for a role of emotions comes from the brief survey that all subjects in the main treatments completed following their decisions. In the survey, subjects were asked to state how pleased or angry they felt about their 1st counterpart's decision, and how much if any sense of obligation they felt to help their 1st counterpart's next partner. The answers, displayed graphically in Appendix Figure B.1, indicate that the strongest emotions, both of anger and of obligation, were felt by cooperators who encountered defectors in their first interaction. Still more suggestive is the fact that among the cooperators who had encountered defectors, average self-reported anger towards the counterpart is significantly higher among those who chose to report than among those who did not choose to report. The reporters also indicated feeling significantly greater obligation towards their counterpart's next partner than did those who did not report.[39,40]

[39] The anger variable is a subject's response to the question: "How did you feel about your first counterpart's decision? Please rate on a scale from 1 = very pleased to 7 = very angry." The obligation to help variable is a subject's response to the question: "Did you feel a sense of obligation to help your first counterpart's next counterpart by sending a report? Please rate on a scale from 1 = did not feel obligated at all, to 7 = felt strongly obligated." Mann-Whitney tests find that the anger levels of the cooperators that reported defectors differ from those of the non-reporters with p = .0446 (two-sided), and that the obligation level of the reporters differs from that of the non-reporters with p = .0009 (two-sided).

[40] As mentioned in footnote 17, one can include subject i's concern for the next counterpart of j, say k, in formal analysis using a social preference model. The Fehr-Schmidt model may not be ideally suited to this, however, since our intuition is that concern for k, if strong enough to influence i's actions, will take the form of a sense of obligation to help k avoid that fate of exploitation by j which i herself has just suffered. If k is a cooperator not yet victimized by another defector, however, including k's payoff in equation (2) actually reduces i's inclination to report, when $\alpha_i \geq \beta_i$, since i's utility would be increased by seeing k suffer at j's hands, bringing k's earnings closer to i's level. Such an envious orientation toward an individual who did one no harm undoubtedly applies to some actors, but it seems inconsistent with the spirit in which Fehr and Schmidt offered their model—a point that has motivated the introduction of more intentions-inclusive social preference models such as Falk and Fischbacher (2006).

6. Conclusion

Numerous experimental studies have found that costly punishment − which can be effective in promoting cooperation − is frequently forthcoming even in the absence of potential strategic benefit to the punisher. The starting point of our paper is the hypothesis that costly *reporting*, which may serve as an indirect form of punishment, might also be frequent in such dilemma situations. More broadly, inclinations to engage in reporting (perhaps closely related to the anthropological and psychological identification of gossip as a human universal) might be critical to the viability of some reputation-based incentive systems. A preference- or emotion-triggered tendency to report at some cost to oneself could be indispensable to social efficiency in situations in which individuals have the option of simply leaving behind undesirable interaction partners without incurring the time, effort or monetary cost to warn others about them.[41]

We conducted a two interaction prisoner's dilemma experiment in which subjects never meet the same partner twice, in which they are wedded to their first choice in both games in the default condition, and in which a subject's first period partner can report the choice of her initial interaction partner to that subject's next partner, who can change his action only if receiving such a report. We varied cost of reporting among four treatments, making it range from 0% to 0.5% to 5% to 10% of the earnings predicted in conventional equilibrium play. The critical feature of our design is that there is no possible selfish material motivation for reporting one's counterpart's choice, just as there is no such motivation for engaging in costly punishment in a one-shot public goods game (Falk *et al*. 2005). Our data show such non-incentivized costly reporting to be common mainly among subjects who choose to cooperate and who encounter a defector: almost

[41] Having information be transmitted as a basis for reputation formation is likely to be critical to decision evolution in most repeated game environments. See, for example, Kamei and Putterman (forthcoming).

59% of those in this situation choose to report despite its cost, whereas less than 17% of those in cooperator-cooperator and defector-defector pairings, and none in defector-cooperator match-ups, choose to report when it is costly.[42]

We used the Fehr-Schmidt model of inequity-averse preferences to provide one example of how a social preference, here concern over earnings "inequity," could motivate costly reporting of defectors by cooperators under plausible beliefs about how next interaction partners would act. Assuming decision-makers focus only on their own payoff and on that of the individual to be reported on, a cooperator's desire to deny a defector the potential reward of unilateral defection in his next interaction is sufficient to motivate costly reporting, according to this model. More generally, combining Fehr and Schmidt's model, their estimate of preference frequencies, and our subjects' self-reported beliefs, generates predictions of the relative frequencies of costly reporting in the four possible cases that rather closely match those in our data. However, we also found evidence that emotions played a role. Specifically, among cooperator-meets-defector pairings, decisions to report that may appear hypothetical when initially made in an added strategy-method treatment are only a third as common as those in the corresponding sequentially designed main treatment. Also, sequential-design reporters on average indicate significantly higher levels of anger towards their first partner than do non-reporters. They also indicate feeling greater obligation to help the next counterpart of their first partner.[43]

[42] Calculations show that the average defector who was reported on earned \$1.25 less in her 2nd interaction than a defector not reported on, hence *ex post* average "induced punishment" exceeded its cost in all treatments.
[43] That many people view defecting against a cooperating partner is grossly unfair is suggested by the fact that more than half of unaffected *third parties* chose to incur a cost to punish a unilateral defector in a laboratory experiment by Fehr and Fischbacher (2004). The decision to report one's counterpart could also be explained by a more general 'demand to express emotion' (see, for example, Grosskopf and López-Vargas, 2014).

Our experiment underscores the fact that reputation formation may involve costs to those sharing or transmitting the information on which reputational knowledge is based, and it illustrates how laboratory experiments can enhance our understanding of the motivations relevant to decisions to bear such costs. We simplified the problem by assuming that whatever information is transmitted is accurate, and that the recipient knows this to be so with certainty.[44] How matters are complicated when the message transmitted is a free choice of the reporter and when the recipient must accordingly decide whether to place trust in it, is a question requiring future research.[45]

[44] Gërxhani *et al*. (2013) likewise impose the constraint that reports are accurate but costly to send. Their finding that many engage in costly reporting resembles ours, although their setting is slightly less complete with respect to ruling out strategic motives of "reputation building" (see above).

[45] Although the possibility that a report is false, and the resulting need for care in assessing it, are doubtless complicating problems, there is nonetheless some evidence of bias toward truth-telling or genuineness of expression, especially when emotion plays a part. A recent example is Fonseca and Peters (2015), who find the preponderance of reports about the trustworthiness of trust game players are truthful, and that the availability of such reports raises efficiency. This parallels the pre-play communication result in trust games found by Ben-Ner and Putterman (2009).

References

Ahn, T.K., Justin Esarey and John T. Scholz, 2009, "Reputation and Cooperation in Voluntary Exchanges: Comparing Local and Central Institutions," *Journal of Politics* 71: 398-413.

Ahn, T.K., Elinor Ostrom, David Schmidt, Robert Shupp and James Walker, 2001, "Cooperation in PD Games: Fear, Greed, and History of Play," *Public Choice* 106: 137-155.

Anderson, Christopher M. and Louis Putterman, 2006, "Do Non-strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism," with Christopher M. Anderson, *Games and Economic Behavior* 54 (1): 1-24.

Ben-Ner, Avner, and Louis Putterman, 2009, "Trust, Communication and Contracts: An Experiment," *Journal of Economic Behavior and Organization* 70: 106 – 121.

Bowles, Samuel and Herbert Gintis, 2004, "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations," *Theoretical Population Biology* 65: 17-28.

_____, 2011, A Cooperative Species: Human Reciprocity and its Evolution. Princeton, NJ: Princeton University Press.

Camerer, Colin, and Richard Thaler, 1995, "Anomalies: Ultimatums, Dictators and Manners," *Journal of Economic Perspectives* 9(2): 209-219.

Camerer, Colin, 2003, *Behavioral Game Theory: Experiments in Strategic Interaction*. New York: Russell Sage Foundation.

Camerer, Colin, and George Loewenstein, 2003, "Behavioral Economics: Past, Present, and Future," in C. Camerer, G. Lowenstein and M. Rabin, *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.

Carpenter, Jeffrey, 2007, "The Demand for Punishment," *Journal of Economic Behavior and Organization* 62: 522-542.

Charness, Gary and Matthew Rabin, 2002, "Understanding Social Preferences With Simple Tests," *Quarterly Journal of Economics* 117(3): 817-869.

Chaudhuri, Ananish, 2011, "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature," *Experimental Economics* 14 (1): 47-83.

Cooper, Russell, Douglas DeJong, Robert Forsythe, and Thomas Ross, 1996, "Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games," *Games and Economic Behavior* 12: 187-218.

Dal Bó, Pedro and Ernesto Dal Bó, 2014, "Do the Right Thing: The Effects of Moral Suasion on Cooperation," *Journal of Public Economics* 117: 28-38.

Dal Bó, Pedro, Andrew Foster and Louis Putterman, 2010, "Institutions and Behavior: Experimental Evidence on the Effects of Democracy," *American Economic Review* 100(5): 2205-2229.

de Quervain, D.J., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Urich Schnyder, Alfred Buck, and Ernst Fehr, 2004, "The Neural Basis of Altruistic Punishment," *Science* 305 (5688): 1254-1258.

Dellarocas, Chrysanthos, 2003, "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* 49(10): 1407-1424.

Dreber A, Rand DG, Fudenberg D, Nowak MA, 2008, "Winners Don't Punish," *Nature* 452: 348-351.

Duersch, Peter and Maroš Servátka, 2009, "Punishment with Uncertain Outcomes in the Prisoner's Dilemma," Working Papers 0485, University of Heidelberg, Department of Economics.

Dunbar, Robin, 2004, "Gossip in Evolutionary Perspective," *Review of General Psychology* 8 (2): 100-110.

Ertan, Arhan, Talbot Page and Louis Putterman, 2009, "Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free-Rider Problem" *European Economic Review* 53: 495-511.

Falk, Armin and Urs Fischbacher, 2006, "A Theory of Reciprocity," *Games and Economic Behavior* 54 (2): 293 – 315.

Falk, Armin, Ernst Fehr and Urs Fischbacher, 2005, "Driving Forces Behind Informal Sanctions," *Econometrica* 73 (6): 2017-2030.

Fehr, Ernst and Colin Camerer, 2007, "Social Neuro-economics: The Neural Circuitry of Social Preferences," *Trends in Cognitive Sciences* 11 (10): 419-427.

Fehr, Ernst and Urs Fischbacher, 2004, "Third-party Punishment and Social Norms," *Evolution and Human Behavior* 25(2): 63 – 87.

Fehr, Ernst and Simon Gächter, 2000, "Cooperation and Punishment in Public Goods Experiments," *American Economic Review* 90 (4): 980-994.

Fehr, Ernst and Klaus M. Schmidt, 1999, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114 (3): 817-868.

Fehr, Ernst and Klaus Schmidt, 2010, "On Inequity Aversion: A Reply to Binmore and Shaked," *Journal of Economic Behavior and Organization* 73: 101-108.

Feinberg, M, Willer, R., Stellar, J., Keltner, D. (2012), "The Virtues of Gossip: Reputational Information Sharing as Prosocial Behavior," *Journal of Personality and Social Psychology*, 102(5): 1015-1030.

Field, Alexander, 2003, *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity*. University of Michigan Press.

Fischbacher, Urs, 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics* 10: 171-178.

Fischbacher, Urs and Simon Gächter, 2010, "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments," American Economic Review 100(1): 541-56.

Fonseca, Miguel and Kim Peters, "Gossip among Trustors Increases Trust and Trustworthiness in the Trust Game," paper presented at Economic Science Association World Meetings, Sydney, Australia, July 2015.

Gächter, Simon and Benedikt Herrmann, 2009, "Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment," *Philosophical Transactions of the Royal Society B* 364(1518): 791-806.

Gërxhani, Klarita, Jordi Brandts and Arthur Schram, 2013, "The Emergence of Employer Information Networks in an Experimental Labor Market," *Social Networks* 35: 541 – 560.

Gintis, Herbert, Samuel Bowles, Robert Boyd and Ernst Fehr, eds., 2005, *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge: MIT Press.

Gregg, Dawn G., and Judy E. Scott, 2006, "The role of reputation systems in reducing online auction fraud," *International Journal of Electronic Commerce* 10.3: 95-120.

Grosskopf, Brit and Kristian López-Vargas, "On the Demand for Expressing Emotions," unpublished paper, University of Exeter and University of Maryland.

Hoffman, Elizabeth, Kevin McCabe and Vernon Smith, 1998, "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry* 36: 335-352.

Jackson, Matthew and Yves Zenou, 2014, "Games on Networks," in Peyton Young and Shmuel Zamir, eds., *Handbook of Game Theory, Vol. 4*. Amsterdam: Elsevier.

Kamei, Kenju, 2014, "Conditional Punishment," *Economics Letters* 124: 199-202.

Kamei, Kenju and Louis Putterman, forthcoming, "Play it Again: Partner Choice, Reputation Building and Learning from Finitely-Repeated Dilemma Games," *Economic Journal* (in press).

Kreps, David, Paul Milgrom, John Roberts and Robert Wilson, 1982, "Rational Cooperation in Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory* 27: 245-252.

Loewenstein, George, 2000, "Emotions in Economic Theory and Economic Behavior," *American Economic Review* 90 (2): 426 – 432.

Nikiforakis, Nikos and Hans-Theo Normann, 2008, "A Comparative Statics Analysis of Punishment in Public Good Experiments," *Experimental Economics* 11(4): 358-369.

Ostrom, Elinor, James Walker and Roy Gardner. 1992, "Covenants with and without a Sword: Self Governance is Possible," *American Political Science Review* 86(2): 404-416.

Palfrey, Thomas and Jeffrey Prisbrey, 1997, "Anomalous Behavior in Public Goods Experiments: How Much and Why?" *American Economic Review* 87(5): 829 – 846.

Pinker, Steven, 2003, *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.

Putterman, Louis, 2014, "When Punishment Supports Cooperation: Insights from Voluntary Contribution Experiments," pp. 17 - 33 in P. Van Lange, B. Rockenbach and T. Yamagishi, eds., *Reward and Punishment in Social Dilemmas*. New York, Oxford University Press.

Rand, David, Hisashi Ohtsuki and Martin Nowak, 2009, "Direct Reciprocity with Costly Punishment: Generous Tit-for-Tat Prevails," *Journal of Theoretical Biology* 256: 45 – 57.

Resnick, P. and R. Zeckhauser, 2002, "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," *Advances in applied microeconomics* 11: 127-157.

Shampanier, Kristina, Nina Mazar, Dan Ariely, 2007, "Zero as a Special Price: The True Value of Free Products," *Marketing Science* 26: 742-757.

Smith, Adam, 1761. *The Theory of Moral Sentiments (2nd ed.)*. Strand and Edinburgh: A. Millar.

Sobel, Joel, 2005, "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, XLIII: 392-436.

Wang, Z., 2010, "Anonymity, social image, and the competition for volunteers: A case study of the online market for reviews," *The BE Journal of Economic Analysis & Policy* 10(1): 1935-1682.

Wilson, David S., 2002, *Darwin's Cathedral: Evolution, Religion, and the Nature of Society*. Chicago: University of Chicago Press.

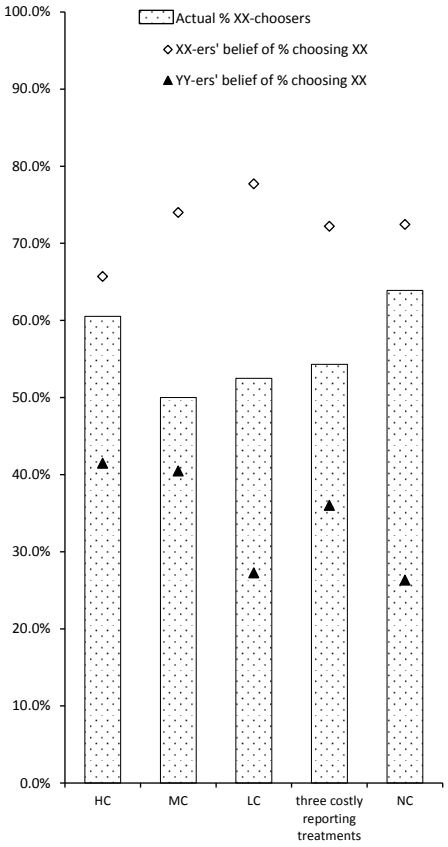
Wilson, Edward O., 2012, *The Social Conquest of Earth*. New York: Liveright.

Yamagishi, T., 1986, "The Provision of a Sanctioning System as a Public Good," *Journal of Personality and Social Psychology* 51(1): 110-116.

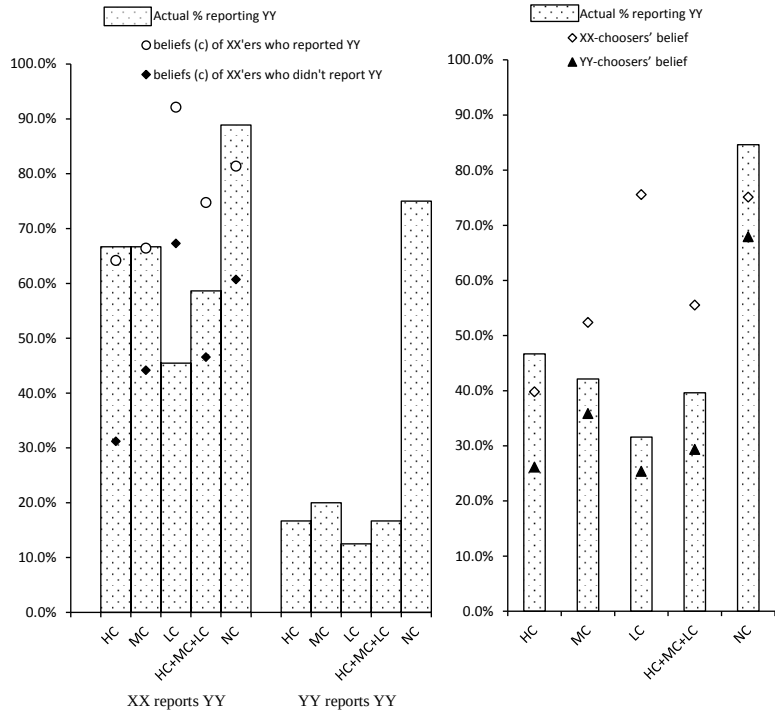
Fig. 1: Payoff Matrix

		Player 2	
		X	Y
Player 1	X	$10, $10	$4, $11
	Y	$11, $4	$5, $5

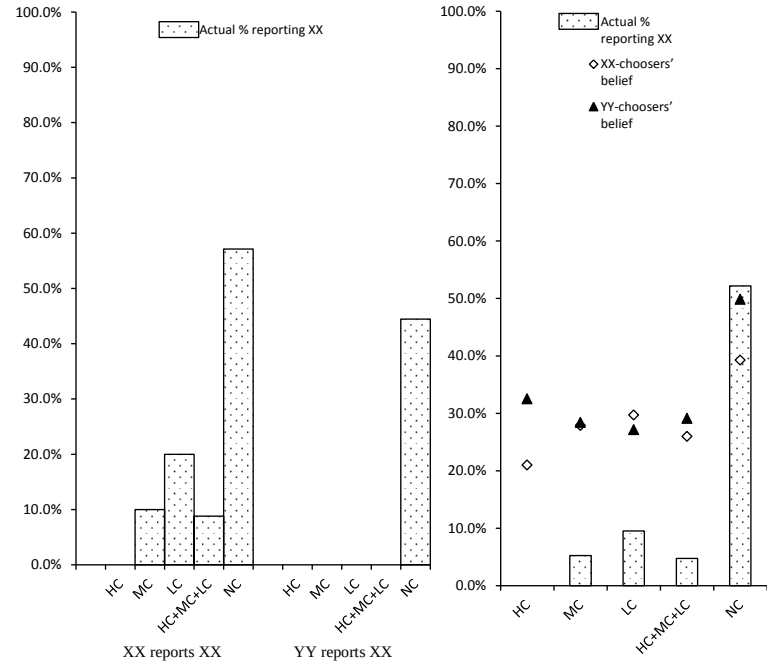
Fig. 2: Choices of XX or YY, Reporting Decisions, and Beliefs



(a) % of choosing XX



(b) % of reporting YY



(c) % of reporting XX

Table 1: Beliefs and Predicted Rationality of Cooperating

	(1) HC	(2) MC	(3) LC	(4) Avg. (1) – (3)	NC
ALL SUBJECTS					
a	0.561	0.572	0.538	0.557	0.558
b	0.256	0.282	0.285	0.274	0.431
c	0.344	0.441	0.517	0.436	0.725
LHS of (1)	0.591	0.896	1.141	0.887	1.545
LHS (1')	2.027	2.509	2.674	2.413	3.950
COOPERATORS					
a(XX)	0.657	0.74	0.777	0.722	0.724
b(XX)	0.21	0.279	0.297	0.26	0.393
c(XX)	0.398	0.524	0.756	0.555	0.751
LHS of (1)	1.016	1.431	2.539	1.620	2.047
LHS of (1')	2.395	3.495	4.847	3.497	4.892
DEFECTORS					
a(YY)	0.415	0.405	0.273	0.36	0.263
b(YY)	0.325	0.284	0.272	0.291	0.499
c(YY)	0.261	0.358	0.254	0.293	0.679
LHS of (1)	0.128	0.508	0.229	0.297	0.916
LHS of (1')	0.937	1.198	0.675	0.925	1.703
"RATIONAL BELIEF"					
r(*a*)	0.605	0.5	0.525	0.543	0.639
r(*b*)	0	0.053	0.095	0.048	0.522
r(*c*)	0.467	0.421	0.316	0.396	0.846
LHS of (1)	1.880	1.341	0.896	1.341	1.881
LHS (1)'	1.880	1.606	1.395	1.601	5.217

Notes: a_i is the proportion of other subjects that subject *i* expects will cooperate (choose XX); b_i the proportion *i* expects will report a first counterpart who chooses XX; c_i the proportion *i* expects will report a first counterpart who chooses YY. As shown in the text, LHS of (1) = $5a_i(c_i - b_i) + c_i$. LHS of (1') = $5a_ic_i + a_ib_i + c_i$. The upper portion shows overall average beliefs by treatment, while the second and third portions from top show average beliefs among XX and YY choosers, respectively. In the bottom portion, r(*a*), etc., represent the beliefs that would have constituted rational expectations, i.e. the actually observed proportions who choose XX, report an XX choice, etc.