

NBER WORKING PAPER SERIES

"DO THE RIGHT THING:
" THE EFFECTS OF MORAL SUASION ON COOPERATION

Ernesto Dal Bó
Pedro Dal Bó

Working Paper 15559
<http://www.nber.org/papers/w15559>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2009

For useful comments and suggestions we thank Rachel Croson, Botond Koszegi, John Morgan, Santiago Oliveros, Parag Pathak, Louis Putterman, Matthew Rabin and Steve Tadelis, as well as conference and seminar participants at Berkeley Econ, Berkeley Haas, Florida State, and UCLA. We thank Pantelis Solomon and Justin Tumlinson for research assistance. We are grateful to the people of Berkeley's XLab for financial and logistic support. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Ernesto Dal Bó and Pedro Dal Bó. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

"Do the Right Thing:" The Effects of Moral Suasion on Cooperation
Ernesto Dal Bó and Pedro Dal Bó
NBER Working Paper No. 15559
December 2009
JEL No. C9,H41

ABSTRACT

The use of moral appeals to affect the behavior of others is pervasive (from the pulpit to ethics classes) but little is known about the effects of moral suasion on behavior. In a series of experiments we study whether moral suasion affects behavior in voluntary contribution games and mechanisms by which behavior is altered. We find that observing a message with a moral standard according to the golden rule or, alternatively, utilitarian philosophy, results in a significant but transitory increase in contributions above the levels observed for subjects that did not receive a message or received a message that advised them to contribute without a moral rationale. When players have the option of punishing each other after the contribution stage the effect of the moral messages on contributions becomes persistent: punishments and moral messages interact to sustain cooperation. We investigate the mechanism through which moral suasion operates and find it to involve both expectation- and preference-shifting effects. These results suggest that the use of moral appeals can be an effective way of promoting cooperation.

Ernesto Dal Bó
University of California, Berkeley
Haas School of Business
545 Student Services Building #1900
Berkeley, CA 94720-1900
and NBER
dalbo@haas.berkeley.edu

Pedro Dal Bó
Department of Economics
Brown University
64 Waterman Street
Providence, RI 02912
and NBER
pdalbo@brown.edu

1 Introduction

The typical assumption in the study of incentives in economics is that people have to be compensated in order to carry out a costly task. However, everyday life abounds in examples where individuals are encouraged not through payments but through appeals of a normative kind. Instances of moral suasion are ubiquitous –they take place in religious ceremonies (avoid sin), political arguments (this policy is right), they are part of educational indoctrination (it is wrong to cheat), marketing strategy (buy fair trade), and the workplace (be a teamplayer). This suggests that there might be room for motivation through moral appeals beyond what money or other forms of compensation can buy.

A large share of empirical and experimental research in economics has been devoted to measuring how extrinsic incentives can be manipulated to affect behavior. We have no equivalent knowledge on the effectiveness of moral suasion. In this paper we report on a series of experiments designed to ascertain the effects of moral suasion on cooperation. We focus on a public good game, or voluntary contribution game, as it offers a clear setting where private and social objectives collide. In such a setting we expose subjects to different messages, some of which contain a moral argument. We then evaluate the effects of messages on subsequent contribution levels.

All four of our experiments involved a voluntary contribution game played through computers. In our first experiment, each session consisted of twenty rounds of a two-person public good game where subjects were randomly rematched after each round. Subjects were given an endowment in each round that they could invest on either a personal account or a joint, “productive” account. Investments in the personal account were retained by the subject. Investments in the joint account were multiplied by 1.4, but divided evenly between the two players of the round, thus yielding an individual net return of only 0.7 per unit invested. The symmetric efficient outcome and Utilitarian optimum is to contribute the entire endowment to the joint account while the unique Nash equilibrium is to contribute zero. Between rounds 10 and 11 subjects saw a randomly chosen message out of a set of possible messages. In the first experiment there were five different messages, including a blank message and two messages with distinct moral content. One stated that moral actions are those that treat others as you would like to be treated. This principle, usually called the “golden rule,” has been present in most culture and religions throughout history (Wattles

1996). The other moral message had a consequentialist, utilitarian root (see Mill 1863). It stated that actions are moral to the extent that they contribute to maximizing collective payoffs. Subjects that could be matched to each other saw the same message.

The first experiment revealed that the moral messages had a positive and significant effect on contributions. Contributions in the pre-message phase were statistically indistinguishable across the five message categories. But the average contributions in the post-message phase of the experiment were higher for the two moral treatments than in the pre-message phase, something that was not true for the other three messages. One of these was a blank message, another was a simple suggestion to contribute that did not involve an explicitly moral backing, and which was included to control for potential demand effects, and the last message stated that in game theory rational and selfish individuals are assumed to maximize their own payoffs.

The effect of the moral appeals was transitory. While contributions in the first few post-message rounds were higher for the moral treatment groups, they were not significantly higher for the last few rounds.

Our second experiment added a punishment stage after the contribution stage in each round, as in Ostrom et al (1992) and Fehr and Gächter (2000). This allowed players to punish low contributors without having to lower their own contributions. We then exposed subjects to one of two messages, either the blank message or the golden rule message. The pre-message rounds displayed higher cooperation levels than in the first experiment, although they continued to display a decreasing trend. The golden rule message triggered a significant increase in contributions. Moreover, in the presence of punishment, the effect of the moral message was persistent. While moral messages alone (in experiment one) and punishments alone (in the pre-message phase in experiment two) did not appear to guarantee high and persistent cooperation, the interactive effect of punishments and a moral appeal did sustain cooperation at fairly high levels.

To summarize, moral suasion has an effect that goes beyond a basic demand effect, and that is persistent in games where players can separately decide on contributions and punishments. A natural question is on the channels through which moral messages may affect behavior. We characterize different mechanisms by relying on a simple model where preferences include a taste for reciprocity as well as an inclination to satisfy a moral imperative.¹

¹On preferences that differ from material payoffs see Andreoni (1990), Levine (1998) and Charness and

A first possibility is that moral messages affect, perhaps temporarily, subjects' preferences, by raising the level of contribution they deem morally right, or by raising the utility weight on meeting that level. We call this a *pure preference* effect. But another possibility is that messages change players' expectations about the behavior of others. To be sure, a voluntary contribution game has a dominant strategy so expectations about the behavior of others is not an issue in principle. But if individuals have a preference for reciprocity, they may want to contribute more if they expect others to do so. In that context, a moral message may simply signal that others will behave better. This may obey to two different reasons. One is that the preferences of others have been affected. The other way in which a subject's beliefs matter is when other players could switch to play a different equilibrium (a purely *belief-based*, or *coordination*, effect). In this paper we present results from two experiments showing that expectations play a role (either in connection with possible changes in the preferences of others or due to pure coordination), and that pure preference effects are also present.

To determine whether expectations matter at all we conducted a modified version of our first experiment where we manipulated subjects' expectations of the probability that other players had seen the same message. We found that the effects of a moral message became weaker when the probability that others had also seen the golden rule message was capped at 50%. This indicates that expectation effects are one way in which moral appeals work. In order to determine whether pure-preference effects are also present, we conducted a final experiment where a subset of players knew that those they were matched with had seen a blank message. We found that among that subset, those receiving a moral message cooperated more than those seeing a blank message. The fact that moral messages have an effect even when holding fixed the message seen by a subject's partner indicates that moral suasion operates partly by shifting the subject's preference over contributions.

The structure of the paper is as follows. The next section describes related literature. Section 3 describes our first experiment and reports its results. Section 4 describes our second experiment and reports its results. Section 5 proposes a framework to distinguish among the types of effects that can be triggered by the messages, and explains how our third and fourth experiments can disentangle them. We then report results from these experiments. Section 6 concludes.

Rabin (2002) among others.

2 Related literature

To our knowledge this paper is the first to report a lab study of the effect of moral discourse on contributions in public good games. Interestingly, in his well known survey Ledyard (1995) mentions that “moral suasion” is one of the forces that may affect behavior in such games but remained unexplored.

Our paper is not however the first to include moral suasion in experiments. Bohm (1972) compares revealed willingness to pay in a public good field experiment across mechanisms, some of which included moral statements. However, his experiment does not allow for a study of the effect of moral statements because the type of mechanism varied together with the presence of those statements.

An earlier antecedent is the work by Schwarz and Orleans (1967). They conducted a field experiment where they interviewed individuals a month before they would file their tax returns. The interviews involved a questionnaire with a subset of questions varying by treatment group. A “sanction” group received some questions asking whether the respondent agreed with different attitudes towards compliance and enforcement approaches. The “conscience” group answered questions mentioning obligation to respect the law, to put the social interest ahead of self-interest, and connections between tax evasion and draft dodging. These authors found significant effects of the “conscience” treatment on tax compliance.

McGraw and Scholz (1991) studied tax compliance after exposing individuals to two different video tapes. One of these mentioned the existence of aggressive legal strategies to minimizing tax exposure while the other mentioned “the importance that Americans place on norms of social responsibility and patriotism, emphasizing how these norms are related to tax compliance.” These authors found no significant effects. A posterior tax compliance field study by Blumenthal, Christian, and Slemrod (2001) in Minnesota also found no significant effects on tax returns of normative messages mentioning fairness. A recent study by Fellner, Sausgruber and Traxler (2009) targeted potential evaders of TV license fees in Austria. They study the effects of sending messages to non-payers reminding them of the need to pay fees, and vary treatments according to various conditions. One of these contained an appeal stating that not paying harmed other households, so paying was a matter of fairness. No significant effects of the normative appeal were found on registration rates. As the authors state, this might be because normative appeals are unlikely to work with people who are

quite inclined to evasion.

Overall, the evidence from field experiments on enforcement tends to indicate that normative appeals are ineffective. Besides sampling issues on selection and size, in a field experiments substantial time may elapse between treatment and action. Subjects may also suspect a selfish manipulation from an authority that is seeking to collect a tax or a fee, and disregard normative appeals. Moreover, the norms of fairness and responsibility that have been invoked in previous experimental work lacked a clear ethical underpinning. These issues raise the question of whether moral suasion is always ineffective.

Our paper also relates to the literature studying the effects of recommendations, without appealing to moral rules, on contributions in public good games. This literature has found limited or no effects of recommendations on contributions (see Marks et al. 1999 and Croson and Marks 2001 for evidence from threshold public good games and Dale and Morgan 2004 for linear public good games).² Interestingly, Dale and Morgan 2004 found that recommendations favoring the top contribution worked less well than recommendations favoring intermediate contributions. The former tended, if anything, to reduce contributions. This provides an interesting contrast with our findings, where effects are positive even when the moral messages recommend the maximum possible contribution level.

Previous literature has shown that communication between subjects can increase contributions in public good games (see Isaac et al. 1985, Isaac and Walker 1988, and Bochet et al. 2006). This paper shows that communication with moral content from the experimenter can affect contributions. It remains to be studied whether communication with moral content from a subject also has an effect.

The results of this paper can be interpreted as capturing the effect of moral framing; see Andreoni (1995) for the effect of framing on public good games. Recent work by Brañas Garza (2006) shows an increase in giving in dictator games where dictators are reminded that “the other players is in your hands,” indicating that a framing that raises personal responsibility for the payoff of others can be effective.

²On the effect of recommendations on coordination games see Van Huyck et al. (1992) and Brandts and MacLeod (1995).

3 Experiment 1: Does moral suasion affect cooperation?

This section covers an experiment that focuses on whether exposure to moral appeals matters for cooperative behavior.

3.1 Experimental design

We conducted 21 experimental sessions at XLAB, UC Berkeley with a total of 320 subjects. The subjects were UC Berkeley students. Subjects interacted exclusively through individual computer terminals. These terminals were separated by lateral partitions that prevented subjects from observing the screens of other subjects' computers. Subjects were paid privately at the end of the session by XLAB personnel. The experimenter's server allocated subjects randomly to groups of eight people. Each player was randomly matched by the server to another person in the group each round. In each round subjects received an endowment of 10 experimental points (or EPs - the exchange rate was 12 EPs for one dollar), and had to decide how much of those to allocate to a personal account and a joint account. Subjects could choose to contribute any number between 0 and 10 up to two decimal points. EPs allocated to the personal account went directly into the person's earnings. EPs going to the joint account got multiplied by an efficiency factor of 1.4, and then divided between the two participants in the interaction. Therefore, the individual return for placing one EP in the joint account was only 0.7 of an EP. It follows that although the Utilitarian optimum and efficient symmetric outcome would be for both players to contribute their whole endowments (leading to payoffs of 14 for each) the Nash equilibrium is for both to contribute zero to the joint account (yielding 10 for each). After each round, players got randomly rematched to another member of their group.

After ten rounds, subjects saw a message in their computers, randomly selected randomly by the server from a set of five possible messages. All subjects in the same group saw the same message. These messages are detailed in Table 1. One was a blank message (henceforth "Blank"), another one contained a suggestion to contribute without moral content (henceforth "Suggestion"), another one expressed the fact that in game theory rational and selfish individuals maximize their own payoffs (henceforth "Nash"), and the other two were

the moral messages. One of these messages expressed that an action is moral if it treats others as you would like to be treated by others (henceforth “Golden Rule”). The other one expressed the Act-Utilitarian standard according to which individual actions are moral if they maximize the sum of all players’ payoffs (henceforth “Utilitarian”).

Two aspects of the moral messages are worth discussing. One is the reason to include two different moral messages. The other one is the precise wording of these messages. The reason to include two different moral messages is that they express very different principles. While the Utilitarian message is consequentialist (the moral tenor of actions depends on their consequences) the Golden Rule principle abstracts from consequences and appeals to a reversibility property (act in a way towards others that you would have others act towards you). As such, this standard is more duty-based, and therefore can be related more closely to the main opponent of consequentialist ethics, namely the deontological Kantian view expressed in the categorical imperative.³ A natural question is whether moral messages matter at all, and if so, whether consequentialist arguments are more or less powerful than duty-based ones.

The precise wording of messages sought to make as clear as possible the messages and their implications. Thus, if no effects were found, one could not argue this had been due to players not fully understanding the normative implications of the messages. Both moral messages as well as the morality-free suggestion to contribute included an added sentence stating “If you were to act according to this rule, you should contribute 10 EPs.”

Players were informed about all details of the game, and about the fact that a message randomly selected by the computer from a set of messages would be shown to them after round 10. At the end of the experiment subjects answered a questionnaire. They were asked to identify the message they had seen, and to provide information about their field of study, gender, SAT scores, and ideology (ranking from 0, most liberal, to 10, most conservative).

3.2 Results

Subjects earned an average of \$23.18, with a minimum of \$18.35 and a maximum of \$29.81. Given that sessions lasted on average less than an hour, the earnings represent a reasonable

³The categorical imperative is to act according to a maxim that one could will to be a universal rule. The golden rule is not equivalent to the Kantian Categorical Imperative (in fact Kant is said to have despised golden-rule - like principles), although it can be derived from it under appropriate restrictions.

hourly rate. A high number of subjects (87%) correctly remembered at the end of the experiment the message that had been shown to their group.

Panel A of Table 2 shows the evolution of contributions to the joint account by round and message in columns (1) and (2). In the first part of the experiment (rounds 1 to 10) the evolution of contributions follows the usual pattern: contributions are substantial at the beginning but decrease as the players gain experience.⁴ It is important to note that there are no significant differences in behavior across groups that ended seeing different messages, consistent with the random assignment of messages.

Did messages affect behavior? From Table 2 we can see that for all messages but the moral messages, contributions were smaller in the second part of the experiment than in the first part. Figure 1 shows average contributions by treatment category for each of the twenty rounds. Figure 2 shows the increase in contributions after the messages were shown. We perform two comparisons. In the first panel of Figure 2 we plot the change the whole pre-message ten rounds to the post-message ten rounds. The second panel of Figure 2 shows the change of the average contribution (by treatment category) from round 10 to round 11.

We aggregate individual contributions at the level of the group and perform Wilcoxon rank-sum tests to statistically compare the contribution increases that occur in the post-message phase. These are reported in Table 2, panel B. We find that the increase in contributions under the moral messages is greater than the increase under the blank message (p-values of 0.093 and 0.016 for Golden Rule and Utilitarian respectively). On the other hand there are no significant differences in the increase of contributions for Nash or Suggestion relative to Blank (p-values of 0.29 and 0.83 for Nash and Suggestion respectively). More importantly, the increase in contributions under the moral messages is greater than under Suggestion (p-values of 0.036 and 0.009 for Golden Rule and Utilitarian respectively). This shows that it is not just the recommendation of a given contribution level that affects behavior, but that the explicitly moral part of the statement has an effect. True, the Suggestion message also generates an increase in contributions in round 11, but such increase is smaller than that induced by the moral messages and the increase erodes faster in the immediately following rounds. This indicates that the overall effect of the moral messages cannot be attributed exclusively to an experimenters' demand effect. The moral messages trigger an additional effect that causes the contributions to jump higher immediately after

⁴For a summary of the literature on public good games see Ledyard (1995).

the message and erode more slowly in the following rounds.

Similar results are obtained if we focus on the change from round 10 to 11 but some of the significance levels are changed. Both the Utilitarian and Golden Rule messages generate significant increases in contributions from round 10 to 11 relative to Blank (p-values of 0.031 and 0.001, respectively –see Table 2, Panel B). While the Utilitarian message seems to have a greater impact than the Golden Rule when we compare part 2 versus 1 (i.e. post-versus pre-message phases) and the opposite happens when we compare round 11 versus 10, these differences are not significant (p-values of 0.401 and 0.4 respectively). The Suggestion message generates a significant increase in contributions from round 10 to 11 –p-value of 0.018, see Table 2, Panel B).⁵ The Golden Rule message generates an increase in contributions from round 10 to 11 that is statistically higher than that of the Suggestion message (p-value of 0.04). In other words, although the suggestion message that is intended to capture demand effects does have an impact on contributions in round 11, two facts are noteworthy. The increase in contributions from round 10 to 11 is higher for the moral messages, and this difference is statistically significant for the Golden Rule message. Second, the increase induced by the Suggestion message erodes immediately. Thus, the long-run effect of messages seen as the impact on the average contribution in the post-message phase relative to the pre-message phase is only significant for the moral messages. This tells us that messages that have an explicitly moral backing have stronger effects than messages that demand contributions without a moral rationale.

The results are also robust to performing statistical tests at the individual level clustering by group. One question to be dealt with in future research is whether the impact of moral messages is due to the fact that the messages are labeled as moral, or to the intrinsic appeal of the principles contained in those statements. In what follows we explore moral suasion in an enriched strategic environment, and later we turn to the issue of the mechanisms behind moral suasion effects.

⁵Since Andreoni and Vesterlund’s (2001) study of altruistic preferences has shown that men are more likely to care about total payoffs and women more likely to care about equality, we could expect the effects of the two moral treatments to differ by gender. We find no significant differences in the response to messages between women and men. However, when comparing part 2 versus part 1, men are significantly more responsive to the Utilitarian message than the Golden Rule message (p-value of 0.074). There is no difference for women. Also regarding the interaction of personal characteristics with the messages, we find little evidence of ideology affecting the response to messages. The exception is that conservative subjects respond to the Nash message by lowering contributions more than liberal subjects from part 2 to 1 (p-value of 0.07).

4 Experiment 2: Moral suasion and punishment

The main take away from our first experiment is that moral appeals can be used to affect cooperation. However, the effects of moral appeals appeared transitory, which could be given at least two interpretations. One interpretation is that moral discourse can be an effective, though short-lived, instrument to promote cooperation. Presumably, new exposure to moral arguments may be required over time. Alternatively, it could be that players, though in principle still willing to cooperate more, eventually start to defect when they observe that not all players abide by the same principles. Such retraction of cooperative behavior may be less common when subjects have the ability to punish players that have been uncooperative. Therefore, it is of interest to study moral suasion in the context of a richer strategic environment to see whether a moral message can trigger a more persistent increase in cooperation. In our second experiment we added in each round a punishment stage after the contribution stage, as in Ostrom et al (1992) and Fehr and Gächter (2000). This allowed players to punish low contributors without having to lower their own contributions.

4.1 Experimental design

The experimental design is as in our first experiment with two modifications. First, we focused on only two messages for reasons of statistical power: Blank and Golden Rule. Second, the stage game was modified to allow subjects to punish their partner after seeing his or her contribution. After players decided their contributions, a screen showed each her own and the other player's contribution and the payoffs to each. Right after a new screen allowed them the possibility to pick how much of a reduction in the other player's payoff to impose. The cost of lowering the other player's payoff in one experimental point was one fourth of an experimental point.

4.2 Results

We conducted 6 experimental sessions at XLAB, UC Berkeley with a total of 136 subjects. The subjects were UC Berkeley students. Subjects earned an average of \$20.71, with a minimum of \$11.93 and a maximum of \$25.45. A high number of subjects (85%) correctly remembered at the end of the experiment the message that had been shown to their group.

Panel A in Table 3 and Figure 3 show the evolution of contributions to the joint account by round and message. Contribution levels before subjects see the messages are greater than in experiment 1, when punishments were not available. This difference is significant (p-value of 0.001). However, it is interesting to note that these high levels of contributions decrease with experience. In fact, the level of contributions in round 10 is significantly smaller than in round 1 (p-value of 0.003). In other words, while punishments help raise the level of contributions in the absence of moral messages, they cannot prevent the erosion of cooperation.

In our new experiment the evolution of contributions before seeing the messages is the same regardless of the message, as it could be expected given the randomization of messages (p-value of 0.5 for rounds 1 to 10 and 0.847 for round 10). Surprisingly this is not always the case for punishments. The groups that ended seeing the moral message appeared to punish more in the first part of the experiment. Columns (3) and (4) in Panel A of Table 3 show the evolution of average punishment by round and treatment category. The difference in average punishment across treatment categories is not statistically significant for the first nine rounds or for the overall average of rounds 1 to 10 (p-value of 0.211) but it is significant in round 10 (p-value of 0.011). Given the controlled nature of the experiment we attribute this imbalance to a random occurrence.

Did messages affect contributions in the presence of punishment? From Table 3 and Figure 4 we see that, aggregating over all rounds before and after the message, the moral message has a positive effect on contributions, while that is not the case for the Blank message. This difference on the impact of the messages is significant (p-value of 0.002 for all rounds –see Table 3, Panel B for the Wilcoxon rank-sum test results). If we compare the change in contributions from round 10 to 11, we also find that Golden Rule has a significantly different effect from the Blank message (p-value of 0.001).

The first graph in Figure 4 shows the differential effect of the moral messages when punishment is possible when we consider all rounds. The increase on the level of contributions caused by the moral message is significantly larger in this experiment than in the first (p-value of 0.021 in a Wilcoxon rank-sum test). This is reported in Table 4 which also shows that the Blank message does not cause a change in behavior when adding punishments (p-values of 0.753 for all rounds and 0.14 for rounds 10 and 11). Interestingly, we do not find a significant difference in the effect of the moral message across experiments if we focus just

on the rounds right before and after the message (p-value of 0.847). This indicates that the main impact of allowing punishments on the effect of the moral message is not on the initial response but on the persistence of this response. In fact, this can be easily seen by comparing the evolution of contributions in the second part of experiments 1 and 2 for the Golden Rule message (compare Panels A in Tables 2 and 3 or Figures 1 and 3). In our first experiment, where punishments were unavailable, contributions decreased markedly with experience after the moral message. This is no longer the case in Experiment 2 which allows for punishments. The moral message interacts with the presence of punishment to increase and sustain higher levels of cooperation.

While it is not central for the issues studied in this paper, it is interesting to broadly examine the connection between moral suasion and punishments. Table 3 and Figure 4 (top right hand panel) show that the moral message increased punishment relative to the Blank message if we aggregate over rounds and compare the pre- and post-message phases. This difference is significant (p-value of 0.002, see Table 3, Panel B).⁶ Given that lower contributions tend to trigger punishment, one would expect the moral message to have two effects on the punishment meted out by a subject: one direct and positive by raising the propensity to punish (holding the contribution of the other player constant), and one indirect and negative by raising the contribution of the other player. The fact that overall the moral message increases both contributions and punishment suggests that the moral message creates a large increase in the propensity to punish for a given level of contribution.⁷

5 How does moral suasion work?

The main conclusion from the first experiment is that exposure to moral appeals affects cooperation rates, and that this effect goes beyond a pure demand effect. Moreover, the

⁶Consistently with the previous literature, we find that subjects tend to punish subjects that contributed less but there are also observations of perverse punishments (subjects that contributed little tend to punish subjects that contributed more than they did). See Fehr and Gächter (2000), Anderson and Putterman (2006) and Carpenter (2007).

⁷Note however that our study is not designed to investigate this assertion in detail. The right way of assessing it would be to study the response of punishment to messages by keeping constant the subject and the combination of contributions by herself and her partner. However, not all subjects will be observed to engage in contributions at the same level after exposure to the message. Those who are may constitute a non-random sample, complicating a precise identification of the effects of moral suasion on the propensity to punish.

second experiment suggests that when players can separately decide on cooperation and punishment, the effects of a moral message on cooperation can be persistent. A natural question is what drives the effects of moral suasion.

One possibility is that the message directly affects preferences. Formally, this amounts to a parametric change in the utility function that represents preferences, which in the space of actions would lead to a shift in each player’s reaction function. A second possibility is that expectations may play a role if subjects have preferences that go beyond maximizing the pecuniary payoffs of the public goods game. For instance, they may not want to contribute if others won’t, but may want to contribute, at least to some extent, if others will.

In this type of setup, players may change their behavior because they believe the message may have affected the preferences of others.⁸ But messages can also have pure coordination effects. These expectation-driven effects require some form of non-selfish preferences, and the fact that moral messages may affect behavior by leveraging those preferences is also intriguing.

In order to clarify the possible role of preferences and expectations, it is useful to lay out a model that encompasses the different possibilities. This model is used exclusively to illustrate the type of effects that we have in mind and is not intended as a representation of the actual preferences of players.

5.1 A simple model

Suppose that each of the two players has a payoff function that contains the payoff from the voluntary contribution game, and an extra term π that captures non-monetary payoffs. In particular, suppose that the utility function of each player i is,

$$\begin{aligned} U(x_i, x_j) &= m - x_i + \alpha(x_i + x_j) + \pi_i, i = 1, 2, \\ \pi_i &= \rho x_i x_j - \gamma(x_i - \chi_i)^2 \end{aligned}$$

where m is the player’s initial endowment (10 tokens in our lab experiment), x_i is the player’s own contribution and x_j is the contribution of the other player. Here α is the voluntary

⁸Similarly, a player may change her behavior because she expects the other player to change his behavior in the belief that her own preferences had changed, and so on with higher order beliefs rooted in the possibility of anyone’s preferences having shifted, even if they have not. In what follows we abstract from these issues and explore a model where there is common knowledge of parameters and rationality.

contribution game's efficiency parameter (in our lab experiments it was 0.7). The term π captures non-monetary payoffs and has two components. The first term $\rho x_i x_j$ captures a reciprocity payoff. This is psychological reward from making a higher contribution x_i when the other player j contributes more. The parameter $\rho \geq 0$ captures the importance of the reciprocity motive. The term $-\gamma(x_i - \chi_i)^2$ captures a 'moral self' payoff; it is a loss function that detracts from the overall payoff depending on how much the player's contribution departs from the moral imperative, or target level $\chi_i \geq 0$ that i believes she should abide by. The parameter γ is a scalar that measures the relative importance of the moral imperative motive relative to reciprocity and pecuniary payoffs.

Given this configuration, allowing parameters to vary arbitrarily could generate very different preferences and equilibria. We are interested in isolating three distinct possibilities with implications for the interpretation of our experimental results. In what follows we fix parameters values at specific levels to facilitate the characterization of the cases of interest although similar cases may arise for a larger set of parameter constellations.

5.1.1 Pure preference effects

Suppose that $\rho = 0$ (i.e. there is no preference for reciprocity), and that $\chi_i > 0, \gamma > 0$ (the player perceives a moral imperative and cares about honoring it). Individual utility for player i becomes,

$$U(x_i, x_j) = m - x_i + \alpha(x_i + x_j) - \gamma(x_i - \chi_i)^2, i = 1, 2.$$

This is a model where individuals care about pecuniary payoff and about approximating a morally ideal behavior. The optimal contribution for player $i = 1, 2$, is

$$x_i^* = \max \left\{ 0, \chi_i - \frac{1 - \alpha}{2\gamma} \right\}.$$

Note that in this game the contribution of each player is independent of that by the other player. The reaction function of player i in the (x_j, x_i) space is a horizontal line. If the payoff function above represents the true preferences, the only way in which the moral message could have an effect in contributions is by affecting a parameter in the utility function. One possibility is that the moral message raises the contribution that is deemed morally desirable

χ_i , or that it raises the relative importance γ of doing the right thing. Note that increases in both parameters will raise equilibrium contributions. These effects exemplify the pure preference effects.

5.1.2 Expectation-driven, but preference-triggered effects

Suppose now that not only the player cares about a moral imperative to contribute a positive amount ($\chi_i > 0, \gamma > 0$) but also that she has a preference for reciprocity $\rho > 0$. The objective function of each player $i, i = 1, 2$, is,

$$U(x_i, x_j) = m - x_i + \alpha(x_i + x_j) + \rho x_i x_j - \gamma(x_i - \chi_i)^2, i = 1, 2.$$

It is easy to see that now each player's contribution is very much dependent on the other player's. The reaction function of player $i, i = 1, 2$, is linear and given by,

$$x_i = \chi_i - \frac{1 - \alpha}{2\gamma} + \frac{\rho}{2\gamma} x_j.$$

Thus, whenever $\chi_i - \frac{1-\alpha}{2\gamma} > 0$ and $\frac{\rho}{2\gamma} < 1$ the reaction function for player i in the space (x_j, x_i) has a positive intercept and a positive slope smaller than one, guaranteeing a single intersection of the reaction functions. That is, equilibrium exists and is unique at a contribution level $x_i = \min \left\{ m, \frac{2\gamma(2\gamma\chi_i + \rho\chi_j) - (1-\alpha)(2\gamma+\rho)}{(2\gamma)^2 - \rho^2} \right\}$.

Under these preferences, the only way in which equilibrium contributions may change is if some parameter changes in one's own utility, or if a player believes a parameter has changed for the other player. Under common knowledge of parameters and rationality, any change in contributions following a change in beliefs must correspond to the moral statement having caused a preference change in at least one of the two players. However, how much a player will adjust his own contribution depends on whether she believes the other player to have seen the same message, as this indicates to what extent, if any, the reaction function of the other player has shifted. This shows that even if players' contributions respond to their expectations about what other players have seen, this may be related to effects ultimately tied to changes in preferences, and not to pure coordination effects. This exemplifies our idea of "expectation-driven, but preference-triggered" effects.

5.1.3 Purely expectation-driven, coordination effects

Lastly, suppose that $\rho > 0$ so each player wishes to reciprocate good or bad behavior but does not feel she needs to match any specific, morally mandated contribution level herself, so $\gamma = 0$. The objective function of player i becomes,

$$U(x_i, x_j) = m - x_i + \alpha(x_i + x_j) + \rho x_i x_j.$$

If each player believes the contribution of the other player to satisfy $x_j > \frac{(1-\alpha)}{\rho}$ then each player will contribute all of her endowment, validating beliefs, leading to an equilibrium with maximal contributions m . But if each player believes the other player's contribution to satisfy $x_j < \frac{(1-\alpha)}{\rho}$ then both players will be in an equilibrium with zero contributions. In this particular case of the model, the players face a coordination game. The reaction functions are step functions with steps at $\frac{(1-\alpha)}{\rho}$. These reaction functions have two stable intersections—one at zero and one at m . In this game, moral messages may have no effect whatsoever in any parameters, and still affect contributions, by directly affecting players' beliefs about which equilibrium is going to be played. This exemplifies the “expectation-driven,” or “pure coordination” effects.⁹

5.2 Experiment 3: Do expectations matter?

In this section we present an experiment that suggests that moral suasion affects behavior in part through changes in the expectations about others.

5.2.1 Experimental design

To determine whether expectations play a role we replicated the experimental design of our first experiment with two modifications. First, we included only the Blank and the Golden Rule messages. Second, we allowed the random message to vary across subjects within the same group of eight. Subjects knew that the probability that any member of their group had seen the same message they had seen was capped at 50%. Thus, the expectations held

⁹Clearly, contributions are not always at the maximum or the minimum levels. Slightly modified models will yield reaction functions that cross at low, but positive contributions, and also at high, but not necessarily maximal, contributions.

by anyone having seen the moral message that any player they were matched with had also seen it were necessarily lower than in the first experiment.

5.2.2 Results

We conducted 6 experimental sessions at XLAB, UC Berkeley with a total of 136 subjects. The subjects were UC Berkeley students. Subjects earned an average of \$23.10, with a minimum of \$18.71 and a maximum of \$27.71. A high number of subjects (91%) correctly remembered at the end of the experiment the message that had been shown to them.

Table 5 and Figure 5 show the evolution of contributions to the joint account by round and message. As before, in the first part of the experiment (rounds 1 to 10) the evolution of contributions follows the usual pattern. Again, it is important to note that there are no significant differences in behavior across subjects that ended seeing different messages, consistent with the random assignment of messages. In this section the unit of observation is the average contribution by group and message. We then compare the contribution rates in the same group by message using the non-parametric Wilcoxon Sign-rank test for matching pairs. We used this test given the lack of independence in behavior of subjects seeing different messages within the same 8 person group

Did messages affect behavior differently than in our first experiment? To answer this question we now focus only on round 11. The reason is that after this round the behavior of subjects is affected by their experience in previous rounds and this may depend on the message seen by other subjects. As different subjects may have played with subjects that saw different messages, behavior after round 11 is less comparable.

From Table 5 and Figures 5 and 6 (bottom left panel), we see that both Blank and Golden Rule result in an increase in average contributions from round 10 to 11 (a restart effect). However, this increase is greater for the Golden Rule message (p-value of 0.037—as reported in Panel B in Table 5). We want to compare the effect of the moral message in this experiment to that in our first experiment. We find that the Blank message does not affect contributions from round 10 to 11 to a different extent in Experiment 3 relative to Experiment 1 (p-value of 0.748; Blank has no effect on contributions in either experiment). However, the effect of the moral message is significantly smaller in Experiment 3 than that observed in our baseline experiment when all subjects saw the same message (p-value of 0.012). This suggests that

expectations play a role in moral suasion and that pure preference effects cannot explain the whole effect of moral messages. Either messages work by changing preferences and this effect is then amplified or mitigated by strategic interaction, or they work exclusively as coordination devices.¹⁰

5.3 Experiment 4: Is there a pure preference effect?

In this section we study whether moral suasion has an effect on behavior that operates directly through preferences, and regardless of expectations on the views or behavior of other players. In this experiment we hold fixed the message seen by a player’s opponent, and compare the players’ behavior depending on whether she has seen a Blank or a moral message. If, holding the other player’s message (and information more generally) fixed, the contribution of a treated player increases under the moral message relative to the Blank one, this will mean that moral suasion affects preferences, and that the role of expectations is complementary. If there is no such increase, this will mean that there are no effects of moral suasion through preferences, and that their effect is purely due to coordination.

5.3.1 Experimental Design

To determine whether moral suasion affects preferences we replicated the experimental design of our first experiment with four modifications. First, we included only the Blank and the Golden Rule messages. Second, the choice of messages and matching of subjects was such that half the subjects saw that their opponent had seen the Blank message. Half of these “informed” subjects saw the Blank message and half saw the Golden Rule message. Subjects knew that if they were informed of their opponent’s message the opponent was not informed about their own message. Third, subjects only participated in one round after the message to eliminate any possibility of repeated interaction effects (which would be problematic to infer effects over preferences).¹¹ Finally, we adjusted the exchange rate to 8 EPs per dollar

¹⁰Again, both possibilities require that subjects care about something other than their own monetary payoffs from the game.

¹¹Under several post-message rounds the following could happen: a subject i that sees the moral message could believe that people tend to imitate behavior and that the person j she is currently matched with may later interact with a person z who has also seen the moral message and who will be matched with i after having encountered j . Not wanting to unfavorably dispose z by sending her a frustrated partner j , i may behave better towards j for reasons other than a change in i ’s preferences. Our design eliminates this possibility.

given the reduction in the number of rounds.

In summary, to test whether moral suasion has an effect through preferences, we compare the behavior of subjects who received a Blank message with those that received the Golden Rule message while holding constant the message seen by those they were playing with (the blank message).

5.3.2 Results

We conducted 10 experimental sessions at XLAB, UC Berkeley with a total of 254 subjects. The subjects were UC Berkeley students. Subjects earned an average of \$19.85, with a minimum of \$15.06 and a maximum of \$23.96. A high number of subjects (79%) correctly remembered at the end of the experiment the message that had been shown to them.

Table 6 and Figure 7 show the evolution of contributions to the joint account by round and message for subjects that ultimately learnt that their partner had seen the Blank message. As before, in the first part of the experiment (rounds 1 to 10) the evolution of contributions follows the usual pattern. Again, it is important to note that there are no significant differences in behavior across subjects that ended seeing different messages, consistent with the random assignment of messages.

From Table 6 and Figures 7 and 8, we see that both Blank and Golden Rule result in an increase in average contributions from round 10 to 11 (there is again a small restart effect). However, this increase is greater for the Golden Rule message (p-value of 0.0004).¹² This suggests that moral suasion affects behavior not only through affecting expectations but it must also have a more direct effect that can be attributed to changing preferences.

6 Conclusion

We report results from four experiments designed to study whether exposure to moral appeals affects cooperative behavior. Moral suasion is ubiquitous in many domains of real life, from family relationships to organizational and political realms. Yet there is a dearth of evidence showing that moral statements can affect behavior. Our paper offers such evidence. However,

¹²In this test the unit of observation is the average contribution by group and message for subjects that saw that their partner in round 11 had seen the Blank message. We then compare for these subjects the contribution rates in the same group by message using the non-parametric Sign-rank test for matching pairs.

our results also indicate that the potential for persistent positive effects depends on aspects of the strategic environment in which moral suasion is used. In our experiment, the interaction of a moral frame and the presence of punishments appears important to sustain cooperation when moral messages or punishments alone could not do so.

An important additional question pertains to the mechanisms through which moral suasion operates. Our design allowed us to identify that moral suasion shifts preferences. But moral suasion also seems to depend on whether players are confident that others have been “treated” as well, highlighting a role for expectations about mutual behavior. When preferences are either purely pecuniary or based on a strictly individual moral imperative those expectation-driven effects cannot arise. Their emergence suggests that moral suasion leverages a pro-social, but also reciprocity-based, aspect of preferences.

The existence of intrinsic preferences such as those based on reciprocity motives is by now well known. However, the fact that intrinsic preferences can be leveraged to affect behavior through relatively cheap methods such as ethical discourse is intriguing, especially when considering that the provision of incentives that target extrinsic motivation is costly. Future work should explore in more detail the variety of settings in which moral framing can be effective at shaping behavior, as well as investigate the interactions between moral suasion and extrinsic incentives.

7 References

- Anderson, C.M. and L. Putterman (2006). “Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism,” *Games and Economic Behavior* 54(1), 1-24.
- Andreoni, J. (1990). “Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving,” *Economic Journal* 100(401), 464-77.
- Andreoni, J. (1995). “Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments,” *Quarterly Journal of Economics* 110(1), 1-21.
- Andreoni, J. and L. Vesterlund (2001). “Which is the Fair Sex? Gender Differences in

- Altruism.” *Quarterly Journal of Economics* 116(1), 293-312.
- Blumenthal, M., C. Christian and J. Slemrod (2001). “Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota,” *National Tax Journal* 54(1), 125-138.
- Bochet, O., T. Page and L. Putterman (2006). “Communication and punishment in voluntary contribution experiments,” *Journal of Economic Behavior & Organization* 60(1), 11-26.
- Bohm, P. (1972). “Estimating Demand for Public Goods: an experiment,” *European Economic Review* 3(1), 111-30.
- Brandts, J. and M.B. MacLeod (1995). “Equilibrium Selection in Experimental Games with Recommended Play,” *Games and Economic Behavior* 11(1), 36-63.
- Carpenter, J.P. (2007). “The demand for punishment,” *Journal of Economic Behavior & Organization* 62(4), 522-42.
- Charness, G. and M. Rabin (2002). “Understanding Social Preferences with Simple Tests,” *Quarterly Journal of Economics* 117(3), 817-870.
- Croson, R. and M. Marks (2001). “The Effect of Recommended Contributions in the Voluntary Provision of Public Goods,” *Economic Inquiry* 39(2), 238-49.
- Dale, D.J. and J. Morgan (2004). “Fairness Equilibria and the Private Provision of Public Goods,” mimeo UC Berkeley.
- Fehr, E. and S. Gächter (2000). “Cooperation and Punishment in Public Goods Experiments,” *American Economic Review* 90(4), 980-94.
- Fellner, G., R. Sausgruber, and C. Traxler (2009). “Legal Threat, Moral Appeal and Social Information: Testing Enforcement Strategies in the Field. Mimeo Max Planck Institute on Collective Goods.
- Fischbacher, U. (2007). “z-Tree: Zurich Toolbox for Ready-made Economic Experiments,” *Experimental Economics* 10(2), 171-178.

- Isaac, R.M., K.F. McCue and C.R. Plott (1985). "Public Good Provision in an Experimental Environment," *Journal of Public Economics* 26(1), 51–74.
- Isaac, R.M. and J.M. Walker (1988). "Communication and free-riding behavior: the voluntary contributions mechanism," *Economic Inquiry* 26, 585–608.
- Ledyard, J. (1995). "Public Goods: A Survey of Experimental Research," pp. 111-94 in John Kagel and Alvin Roth, eds., *Handbook of Experimental Economics*. Princeton: Princeton University Press.
- Levine, D.K. (1998). "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics* 1(3), 593-622.
- Marks, M.B., D.E. Schansberg, and R.T.A. Croson (1999). "Using Suggested Contributions in Fundraising for Public Good: an Experimental Investigation of the Provision Point Mechanism," *Nonprofit Management & Leadership* 9(4), 369-384.
- McGraw, K. and J. Scholz (1991). "Appeals to Civic Virtue Versus Attention to Self-Interest: Effects on Tax Compliance," *Law and Society Review* 25(3), 471-498.
- Mill, J.S. (1863). *Utilitarianism*. In *On liberty and other essays* (1991), Oxford University Press.
- Ostrom, E., J. Walker, and R. Gardner (1992). "Covenants With and Without a Sword: Self-Governance is Possible," *American Political Science Review* 86(2), 404–17.
- Schwarz, R. and S. Orleans (1967). "On Legal Sanctions," *University of Chicago Law Review* 34, 274-300.
- Van Huyck, J.B., A.B. Gillette and R.C. Battalio (1992). "Credible Assignments In Coordination Games," *Games and Economic Behavior* 4(4), 606-26.
- Wattles, J. (1996). "The Golden Rule," Oxford University Press.

Table 1: Treatment Messages

Name	Message
1 Blank	BLANK MESSAGE
2 Nash	Please read this message carefully: The assumption of game theory is that rational and self-regarding individuals will maximize their own payoffs. If you were to act accordingly, you would allocate 0 to the joint account.
3 Golden rule	Please read this message carefully: An action of yours is moral if it treats others the way you would like others to treat you. If you were to act accordingly, you would allocate 10 to the joint account.
4 Utilitarian	Please read this message carefully: An action of yours is moral if it maximizes the sum of everyone's payoffs. If you were to act accordingly, you would allocate 10 to the joint account.
5 Suggestion	Please read this message carefully: You could consider allocating all your endowment to the joint account. If you were to act accordingly, you would allocate 10 to the joint account.

Table 2: Does moral suasion affect cooperation? - Experiment 1

Panel A: Contributions by Period and Message						
Round	Message					
	Blank	Nash	Golden Rule	Utilitarian	Suggestion	
	(1)	(2)	(3)	(4)	(5)	
1	3.25	3.55	3.59	4.38	3.22	
2	3.02	2.98	3.01	3.70	3.38	
3	2.60	2.82	3.09	3.46	3.06	
4	2.28	2.62	2.73	3.07	2.55	
5	2.05	2.68	1.91	3.01	2.56	
6	1.85	2.16	1.47	2.96	2.72	
7	2.17	2.52	1.53	2.97	2.19	
8	1.87	2.06	1.10	2.68	2.62	
9	1.82	1.81	1.18	2.47	2.16	
10	1.64	1.66	1.11	2.47	1.77	
11	2.18	1.58	4.38	4.97	3.70	
12	1.52	1.78	3.43	4.17	2.44	
13	1.57	1.58	2.58	4.05	2.06	
14	1.62	1.42	2.51	3.86	1.86	
15	1.26	1.31	1.71	3.69	1.74	
16	1.25	1.35	1.47	3.08	1.50	
17	1.60	1.32	1.44	3.22	1.17	
18	1.18	0.97	1.13	2.98	1.04	
19	1.11	0.66	1.00	2.45	0.95	
20	1.02	1.09	1.23	2.45	1.39	
Average contributions and differences						
Round 11 - Round 10	0.54	-0.09	3.27	2.50	1.93	
Part 1 (pre-message)	2.25	2.49	2.07	3.12	2.62	
Part 2 (post-message)	1.43	1.31	2.09	3.49	1.78	
Part 2 - Part 1	-0.82	-1.18	0.01	0.37	-0.84	
Number of subjects	64	64	64	64	64	
Panel B: Non-parametric p-values (Wilcoxon rank-sum tests)						
Part 2 versus Part 1						
	Nash	Golden Rule	Utilitarian	Suggestion		
Blank	0.294	0.093	0.016	0.834		
Nash		0.016	0.002	0.208		
Golden Rule			0.401	0.036		
Utilitarian				0.009		
Round 11 versus Round 10						
	Nash	Golden Rule	Utilitarian	Suggestion		
Blank	0.092	0.001	0.031	0.018		
Nash		0.001	0.013	0.006		
Golden Rule			0.400	0.040		
Utilitarian				0.371		

Note: we test the hypothesis that the change in contributions from part 1 to part 2 or from round 10 to 11 for groups in different treatment categories stem from different distributions, treating the change in the average contribution of each 8-person group as a single observation.

Table 3: The effects of moral suasion when punishment is available - Experiment 2

Panel A: Behavior by Period and Message				
Round	Contributions		Punishments	
	Blank (1)	Golden Rule (2)	Blank (3)	Golden Rule (4)
1	4.75	4.46	0.99	1.39
2	4.49	4.60	0.95	1.38
3	3.83	4.57	1.38	1.51
4	3.93	4.28	0.80	1.64
5	3.67	4.24	1.17	1.52
6	3.82	3.99	1.17	1.85
7	3.25	3.41	1.03	2.58
8	2.93	3.44	1.51	1.80
9	2.85	3.13	1.08	2.62
10	2.88	3.08	1.10	2.87
11	3.03	6.19	1.57	2.58
12	2.74	5.50	1.07	2.25
13	2.73	5.41	1.00	1.96
14	2.76	5.17	1.23	2.40
15	2.47	5.33	1.05	2.89
16	2.76	5.56	0.97	2.64
17	2.80	5.44	1.13	3.13
18	2.41	4.93	1.19	3.45
19	2.51	4.92	0.88	3.50
20	2.61	5.01	0.80	2.93
Average contributions and differences				
Round 11 - Round 10	0.15	3.11	0.47	-0.29
Part 1 (pre-message)	3.64	3.92	1.12	1.91
Part 2 (post-message)	2.68	5.35	1.09	2.77
Part 2 - Part 1	-0.96	1.43	-0.03	0.86
Number of subjects	64	72	64	72
Panel B: Non-parametric p-values (Wilcoxon rank-sum tests)				
Blank-Golden Rule	Contributions		Punishments	
	Part 1 vs 2	Round 10 vs 11	Part 1 vs 2	Round 10 vs 11
	0.002	0.001	0.002	0.067

Note: in Panel B we test the hypothesis that the change in contributions and punishments from part 1 to part 2 or from round 10 to 11 for groups in different treatment categories stem from different distributions, treating the change in the average contribution of each 8-person group as a single observation.

Table 4: Difference between Experiments 1 and 2: Are contributions different?

Non-parametric p-values (Wilcoxon rank-sum tests)

	Contributions	
	Part 1 vs. 2	Round 10 vs. 11
Blank	0.753	0.14
Golden Rule	0.021	0.847

Note: we test the hypothesis that the change in contributions from part 1 to part 2 or from round 10 to 11 for groups in different experiments within treatment categories stem from different distributions, treating the change in the average contribution of each 8-person group as a single observation.

Table 5: Do expectations play a role? - Experiment 3

Panel A: Contributions by Period and Message		
Round	Message	
	Blank (1)	Golden Rule (2)
1	3.17	3.39
2	3.30	3.13
3	3.35	3.13
4	3.12	2.33
5	2.99	2.29
6	2.78	2.01
7	2.80	1.70
8	1.84	2.44
9	1.71	1.71
10	1.86	1.52
11	2.20	3.05
12	2.19	2.26
13	1.75	2.38
14	1.74	2.21
15	1.53	1.71
16	1.74	1.81
17	1.60	1.58
18	1.34	1.35
19	1.42	1.09
20	1.46	1.14
Average contributions and differences		
Round 11 - Round 10	0.35	1.52
Part 1 (pre-message)	2.69	2.36
Part 2 (post-message)	1.70	1.86
Part 2 - Part 1	-0.99	-0.51
Number of subjects	69	67

Panel B: Non-parametric matched pairs p-values	
	Round 10 vs 11
Blank-Golden Rule	0.037

Note: in Panel B we compare contributions between treatments using a matched pairs test. Each observation is the difference in the change in contributions from round 10 to 11 between subjects that saw the Golden Rule and subjects that saw the Blank message in a group.

Table 6: Are There Pure Preference Effects? - Experiment 4

Contributions by Period and Message for Subjects

Who Know that Partner saw a Blank Message

Round	Message	
	Blank (1)	Golden Rule (2)
1	2.92	3.08
2	2.89	2.67
3	2.11	2.44
4	2.03	2.28
5	2.09	2.08
6	1.62	2.12
7	1.38	1.83
8	1.44	1.26
9	1.64	1.31
10	1.39	1.17
11	1.65	3.27
Round 11 - Round 10	0.26	2.10
Number of subjects	66	66

Figure 1: Contributions by Round and Message – Experiment 1

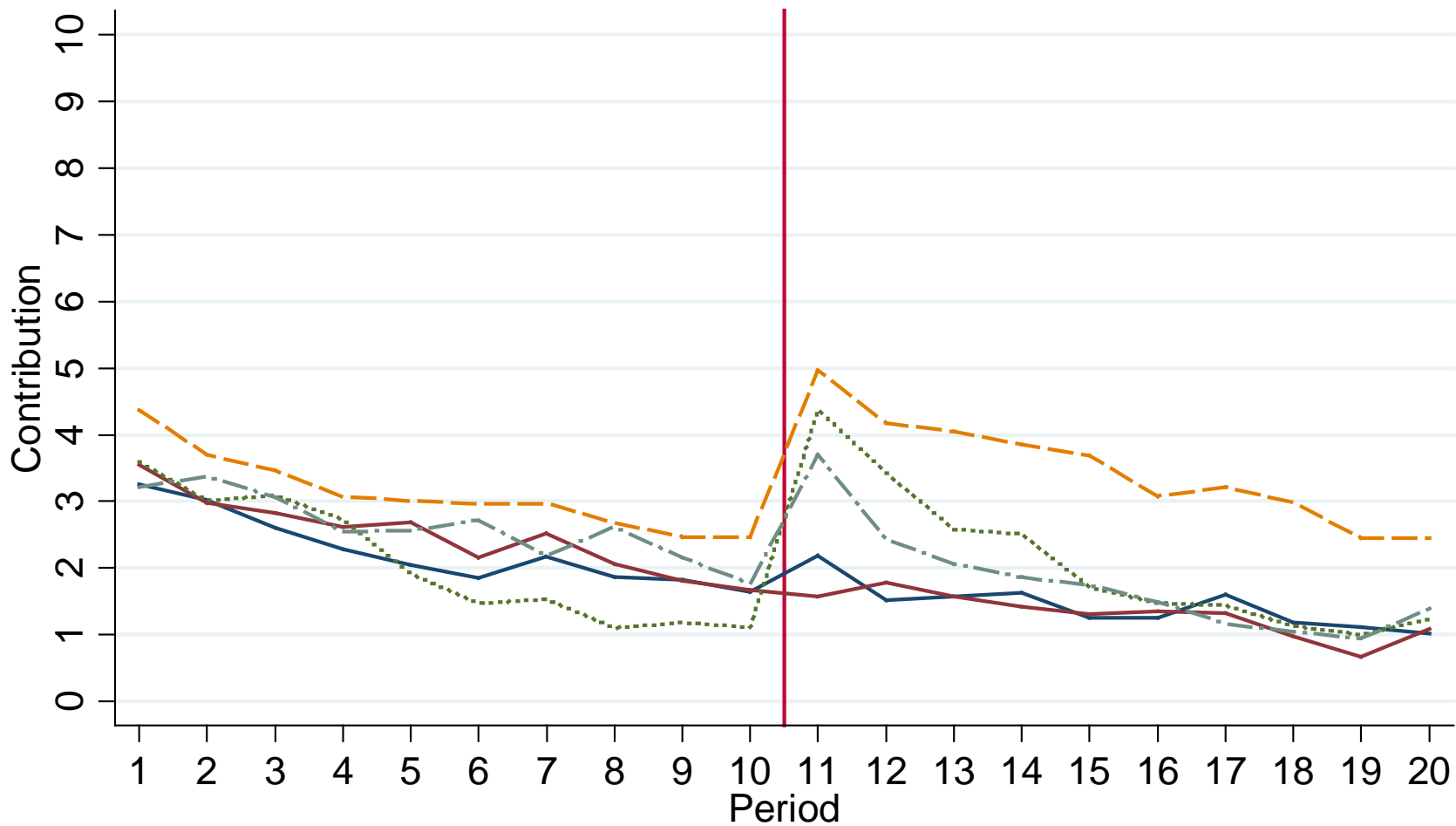


Figure 2: Change in Contributions after Message – Experiment 1

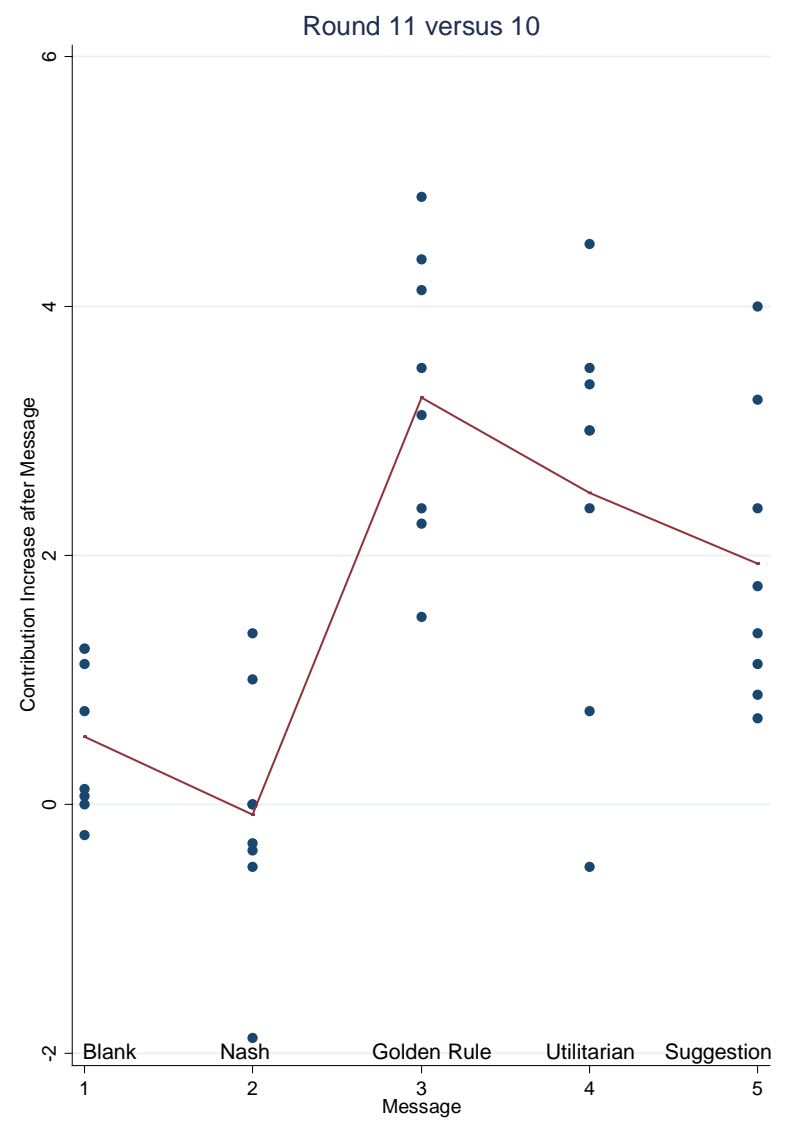
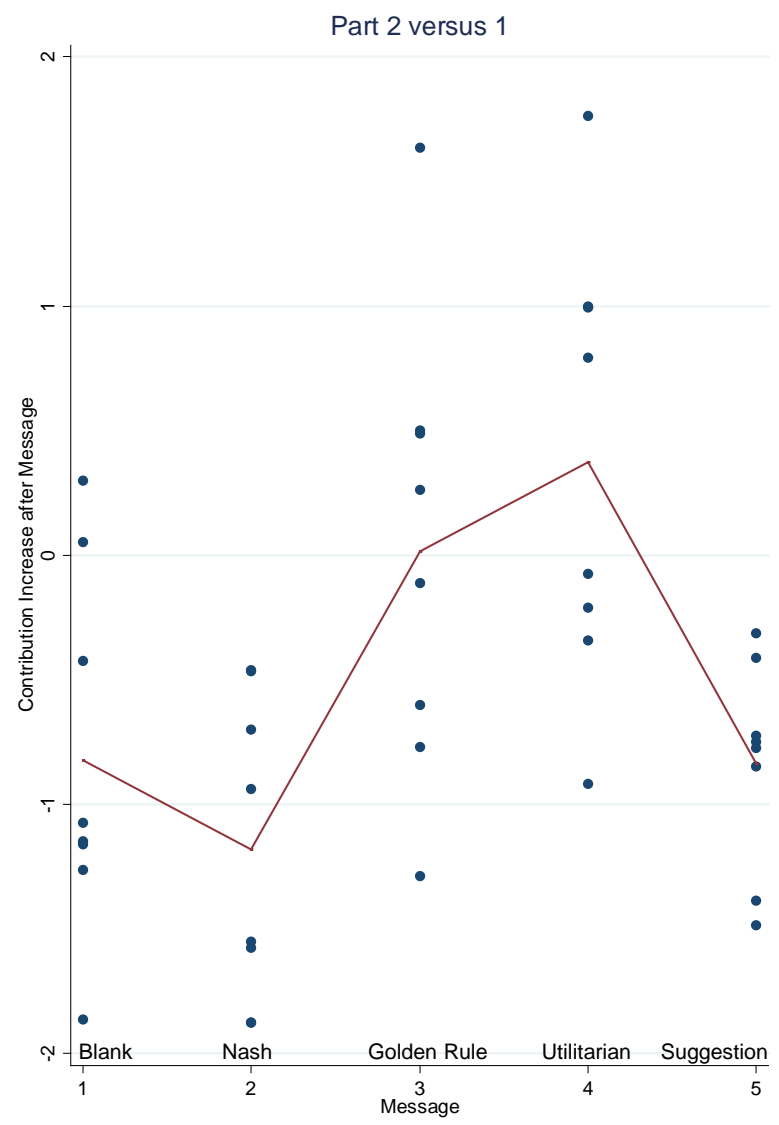


Figure 3: Contributions by Round and Message – Experiment 2

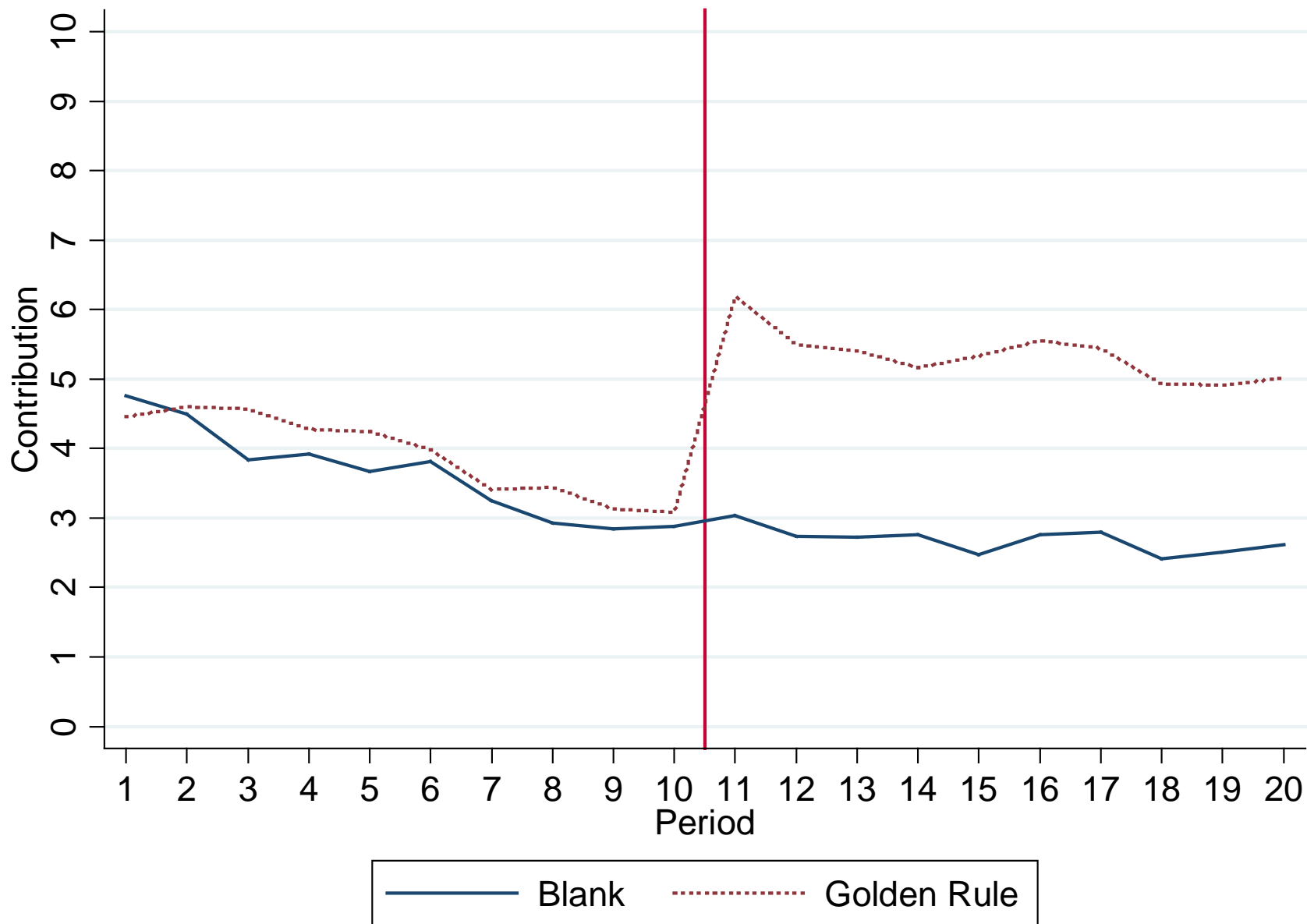
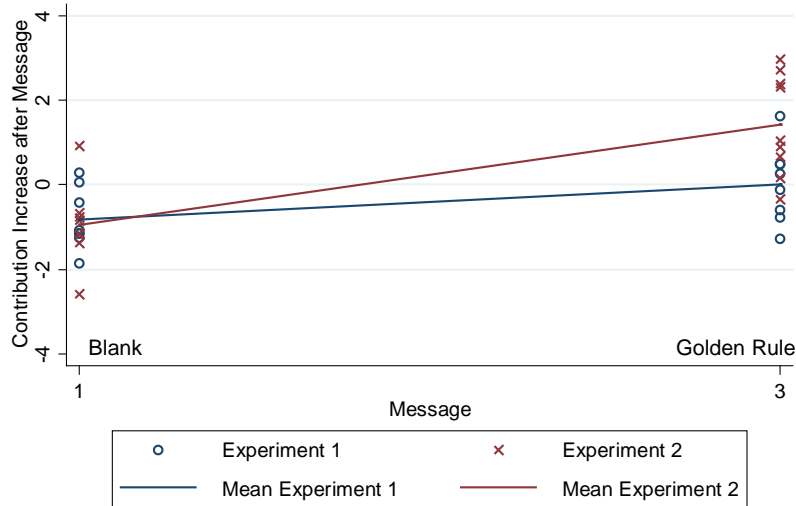


Figure 4: Change in Behavior after Message – Experiment 2

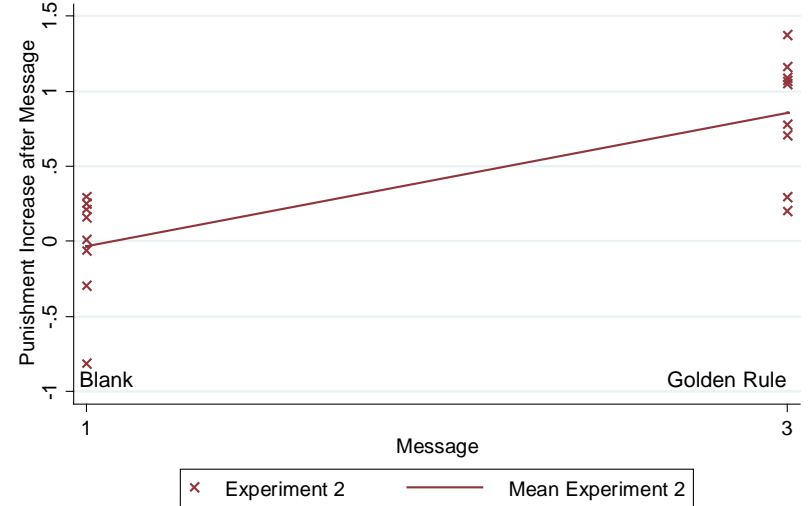
Contributions

Part 2 versus 1

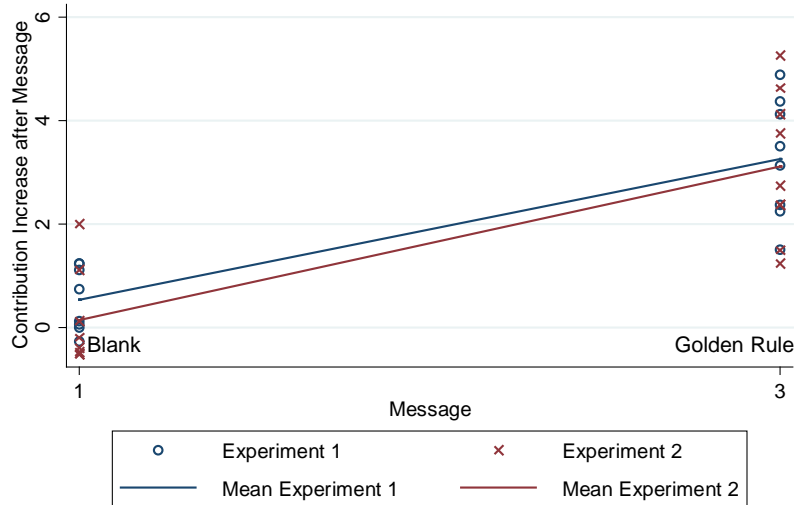


Punishments

Part 2 versus 1



Round 11 versus 10



Round 11 versus 10

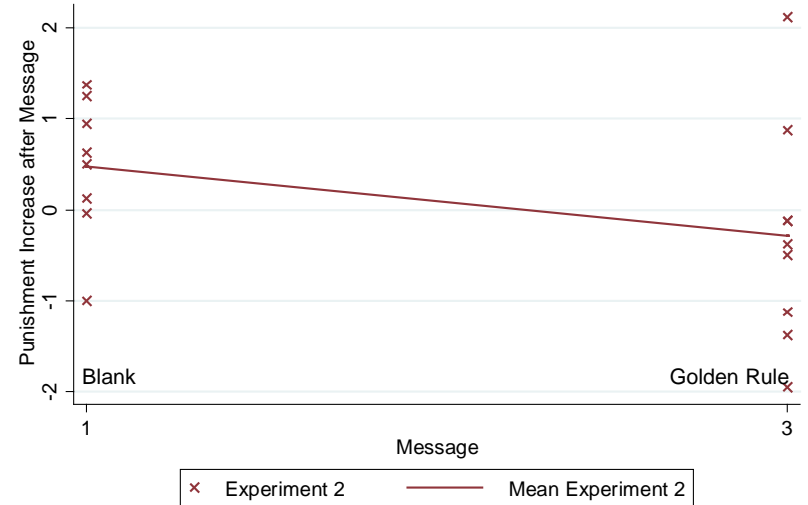


Figure 5: Contributions by Round and Message – Experiment 3

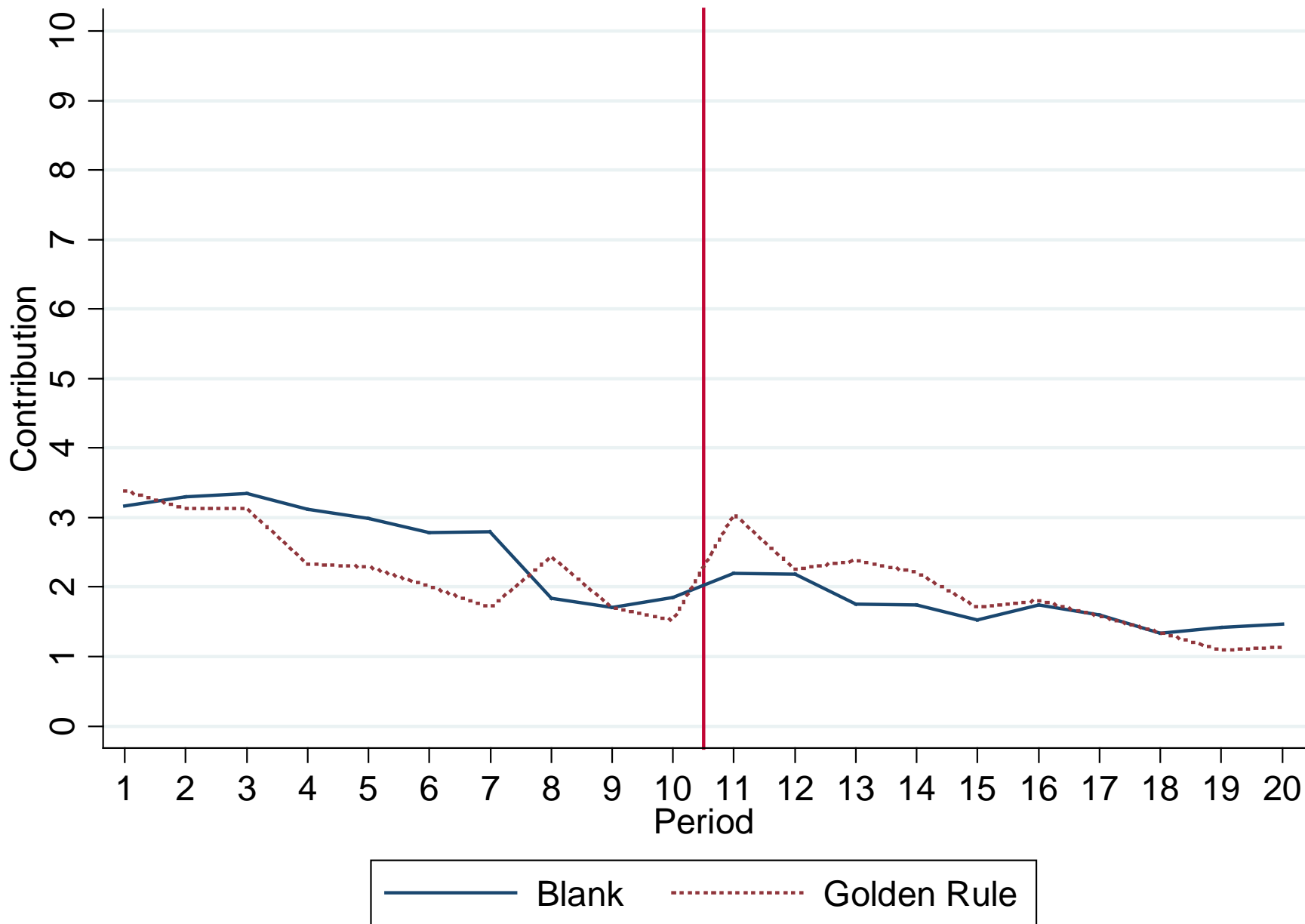


Figure 6: Change in Contributions after Message – Experiment 3

Round 11 versus 10

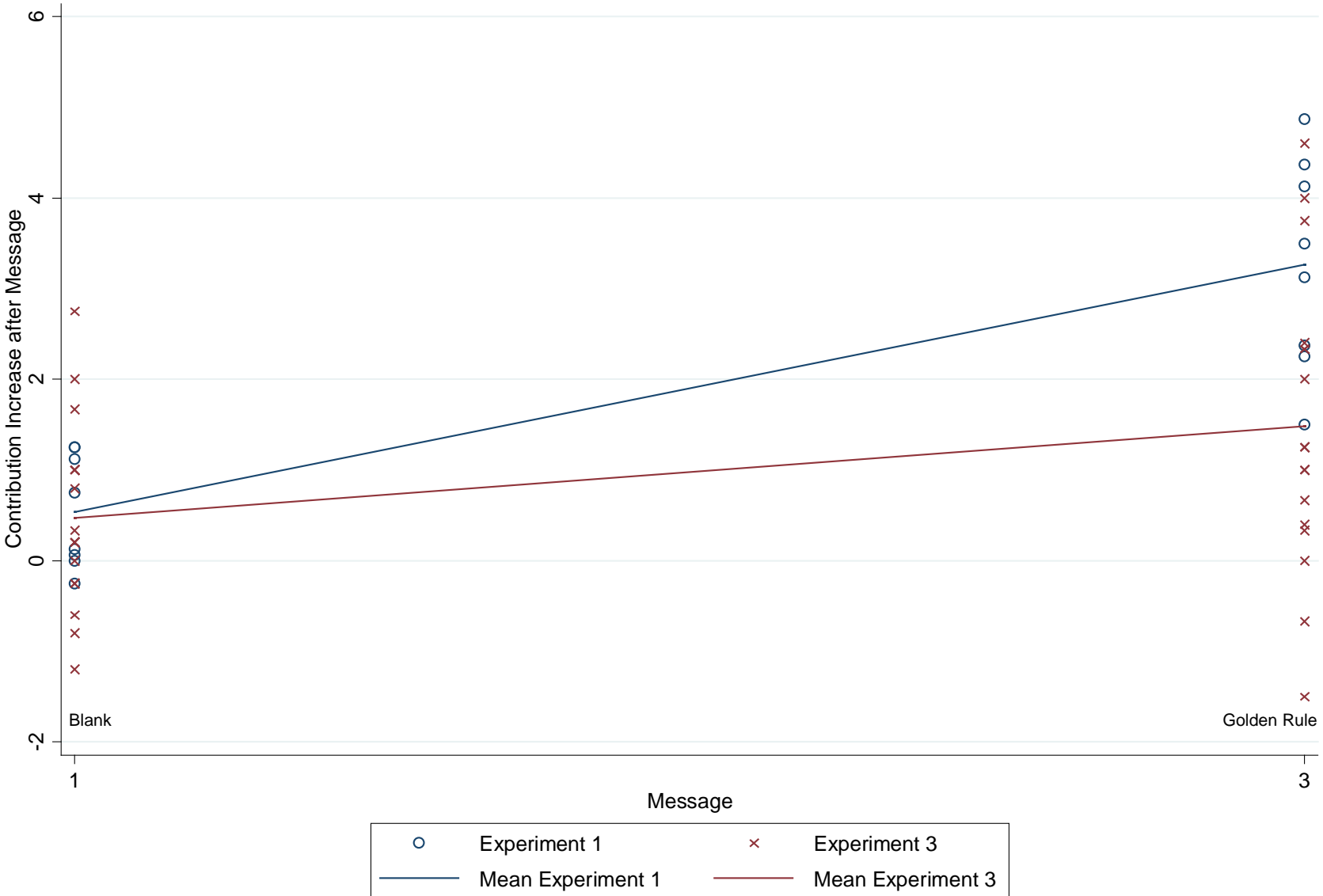


Figure 7: Contributions by Round and Message – Experiment 4

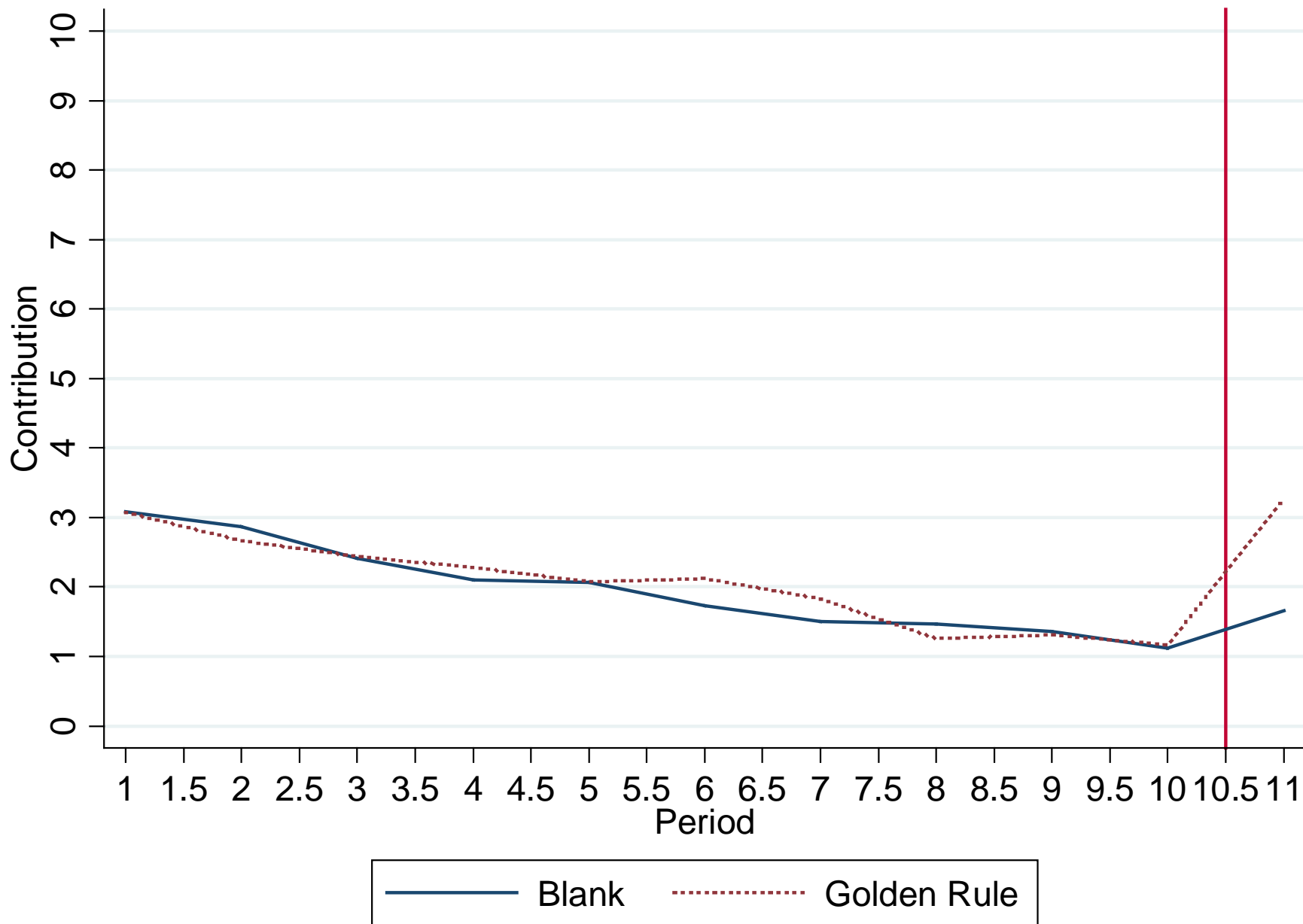


Figure 8: Change in Contributions after Message when Knowing that Other Saw Blank Message – Experiment 4

Round 11 versus 10

