

# “Do the Right Thing:” The Effects of Moral Suasion on Cooperation

Ernesto Dal Bó

Pedro Dal Bó

UC Berkeley & NBER

Brown University & NBER\*

August 20, 2010

## 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 the 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.

*JEL codes: C9, H41.*

*Keywords: moral suasion, morality, cooperation, public goods, ethics.*

---

\*For useful comments and suggestions we thank Anna Aizer, Alberto Alesina, Rachel Croson, Erik Eyster, Botond Koszegi, John Morgan, Santiago Oliveros, Parag Pathak, Louis Putterman, Matthew Rabin and Steve Tadelis, as well as seminar and conference participants at Berkeley, FSU, Harvard/MIT, Brown, UCLA and Ecole Polytechnique. We thank Pantelis Solomon and Justin Tumlinson for research assistance as well as Berkeley’s XLab for financial and logistic support.

# 1 Introduction

While economics pays great attention to the use of material incentives to shape behavior, everyday life abounds in examples where individuals are encouraged not through material incentives 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.

Much empirical and experimental research in economics has been devoted to measuring how material 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 expose subjects to different messages, some of which contain a moral argument. We then evaluate the effects of messages on subsequent contribution levels in a public goods game. We focus on this game to explore not only whether moral appeals affect behavior but also how moral suasion depends on aspects of strategic interaction.

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 five possible messages, including 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 cultures 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.

The remaining three messages were as follows. One was a blank message (control). 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.<sup>1</sup> The last message stated that in game theory rational and selfish individuals are assumed to maximize their own payoffs. All subjects in the same group 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 messages. 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. 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 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 sensitive to the strategic environment. The effect becomes persistent in games where players can separately decide on contributions and punishments. A natural question concerns the channels through which moral suasion operates, and whether moral suasion can be amplified in social contexts.

A first possible channel through which moral suasion may operate is by affecting, perhaps temporarily, subjects' preferences. A moral message may raise the level of contribution

---

<sup>1</sup>An experimenters' demand effect arises when subjects' behavior is affected by what they think the experimenter desires.

subjects deem morally right regardless of what others do, or by raising the utility weight on meeting that level. This effect entails a shift in a player’s best response function, which we refer to as a preference effect.

Another possibility is that messages change players’ expectations about others. For example, a moral message that is commonly observed may signal that others will increase their contributions, and thus affect the behavior of subjects with preferences for reciprocity.<sup>2</sup> In this way, moral messages may affect behavior through changes in expectations, which we refer to as an expectation effect.

We find that moral suasion triggers both preference and expectation effects. 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 the expectation effect is one way in which moral appeals work. This implies that a social amplification effect is present: moral suasion is stronger when we are more confident that other players are getting the same message. In order to determine whether a preference effect also operates, we conducted another experiment where a subset of players knew that those with whom they were matched 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 (blank) message seen by a subject’s partner indicates that moral suasion operates partly by shifting the subject’s preference over contributions.

## 2 Related literature

To our knowledge this paper is the first to report a laboratory study of the effect of moral suasion 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 remains unexplored.

Our paper is not however the first to include moral suasion in experiments. Bohm (1972)

---

<sup>2</sup>On preferences that differ from material payoffs see Andreoni (1990), Levine (1998) and Charness and Rabin (2002) among others.

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.

Field experiments on the effect of normative appeals on tax evasion have found no effects (see McGraw and Scholz 1991, Blumenthal et al. 2001, and Fellner et al. 2009). The exception is Schwarz and Orleans (1967), but their design confounds normative appeals with other factors that can affect compliance. Besides differences in sample selection and size, substantial time may elapse between treatment and action in field experiments. 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 nor incentives, 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).<sup>3</sup> 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). It remains to be studied whether communication with moral content from a subject also has an effect, while our paper shows that communication with moral content from the experimenter can affect contributions.

The results of our paper can be interpreted as capturing the effect of moral framing; see Andreoni (1995) for the effect of framing on public good games.<sup>4</sup>

---

<sup>3</sup>On the effect of recommendations on coordination games see Van Huyck et al. (1992) and Brandts and MacLeod (1995). There is also a literature on how laws can express expected rules of behavior and affect behavior, see Cooter (1995), Bohnet and Cooter (2005) and Galbiati and Vertova (2008).

<sup>4</sup>Framing plays a role in dictator games, too. Brañas Garza (2006) shows an increase in giving in dictator

### 3 Experiment 1: Does moral suasion affect cooperation?

This section covers an experiment that shows that exposure to moral appeals affects 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 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 (hence-

---

games where dictators are reminded that "the other player is in your hands," indicating that a framing that raises personal responsibility for the payoff of others can be effective. Communication between subjects in dictator games may induce similar effects (Andreoni and Rao 2009).

forth “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.<sup>5</sup> 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).

---

<sup>5</sup>The 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.

## 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 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 and Figure 1 show the evolution of contributions to the joint account by round and message. 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.<sup>6</sup> 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 2 shows in two ways the increase in contributions after the messages were displayed. In the first panel of Figure 2 we plot the change in average contributions from the first 10 rounds (pre-message) to the last 10 rounds (post-message). The second panel of Figure 2 shows the change in average contributions from round 10 to round 11.

To statistically compare the contribution increases that occur in the post-message phase we aggregate individual contributions at the level of the group and perform Wilcoxon rank-sum tests. These are reported in Table 2, panel B. All our p-values are obtained by permutation of treatment status (with 10,000 repetitions), and express the fraction of permutations yielding a larger (or smaller, depending on the hypothesis) change than the one observed in the data -thus, these are one-tailed tests.<sup>7</sup>

We focus first on the change in average contributions from the first 10 rounds (pre-message) to the last 10 rounds (post-message). We find that the increase in contributions under the moral messages is greater than the increase under the blank message (p-values of 0.053 and 0.007 for Golden Rule and Utilitarian respectively). On the other hand there are no significant differences in terms of a decrease of contributions under Nash or an increase under Suggestion relative to Blank (p-values of 0.16 and 0.46 respectively). More importantly, the increase in contributions under the moral messages is greater than under Suggestion (p-

---

<sup>6</sup>For a summary of the literature on public good games see Ledyard (1995).

<sup>7</sup>The results are robust to performing statistical tests at the individual level clustering by group.

values of 0.019 and 0.004 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. This indicates that the overall effect of the moral messages cannot be attributed exclusively to an experimenters' demand effect.

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 0.0001 and 0.015, respectively –see Table 2, Panel B). The Suggestion message generates a significant increase in contributions from round 10 to 11 relative to Blank (p-value of 0.009). 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 0.02). 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.

Regarding the difference between the two moral messages, we find that 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, but these differences are not significant (p-values of 0.22 and 0.8 respectively).<sup>8</sup>

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.

---

<sup>8</sup>The effect of moral messages on average contributions is due to an increase in contributions of both those who were already contributing and those who were not contributing before seeing the message.

## 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. It could also 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 to lower the other player's payoff. 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 0.0003). 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  $<0.0001$ ). 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 0.28 for rounds 1 to 10 and 0.44 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 0.12) but it is significant in round 10 (p-value of 0.005). 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 0.001 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 0.0002).

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 0.012 in a Wilcoxon rank-sum test). This is reported in Table 4 which also shows that adding punishments did not change the effect of the Blank message (p-values of 0.636 for all rounds and 0.9301 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 0.594). 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 cooperation and sustain it at higher levels.

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 shows that the moral message significantly increased punishment relative to the Blank message if we aggregate over rounds and compare the pre- and post-message phases (p-value of 0.0004). However, if we focus on rounds 10 and 11 we find the opposite.<sup>9</sup> 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 by changing 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 reduction in punishment in groups that received the moral message relative to the Blank message can simply be explained by the increase in contributions in the former. However, the fact that moral messages increase both contributions and punishment when we consider all rounds suggests the moral message may increase the propensity to punish for a given level of contribution by the other.<sup>10</sup>

---

<sup>9</sup>Consistently 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).

<sup>10</sup>Note however that our study is not designed to investigate this assertion in detail. A way to assess 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.

## 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 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 moral suasion may affect subjects' preferences by raising the level of contribution subjects deem morally right or by raising the utility weight on meeting that level. This preference effect would result in a shift in a player's best response function.

A second possibility is that moral suasion changes players' expectations about others. 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 that is commonly observed may signal that others will contribute more, and affect behavior.<sup>11</sup> This expectation effect highlights a "social" aspect of moral suasion, namely that the effectiveness of moral appeals could depend on the fact that individuals are interacting with others who are also receiving the moral appeal.

We use two experiments to determine whether expectation and/or preference effects are present.

### 5.1 Experiment 3: Do expectations matter?

This section covers an experiment that shows that moral suasion affects behavior in part through changes in the expectations about others.

#### 5.1.1 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

---

<sup>11</sup>Moral suasion may also affect second order beliefs, which can also affect behavior if subjects' preferences depend on these beliefs (see Geanakoplos et al. 1989 on psychological games). We do not study in this paper whether it is first or higher order beliefs which matter for the expectation effect.

had seen the same message they had seen was capped at 50%. Since subjects in the same group could see different messages, the expectations held by anyone having seen the moral message that any peer had also seen it were necessarily lower than in the first experiment. Therefore, if expectations were important to moral suasion we would expect the effects to be weaker in this experiment than in our first one.

### 5.1.2 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.

Did messages affect behavior differently than in our first experiment? To answer this question we 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, rounds after 11 are less comparable.

From Table 5 and Figures 5 and 6, 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 0.01).<sup>12</sup>

More importantly, we compare the effect of the moral message in this experiment to that in our first experiment. 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 0.0036) while there are no differences for the Blank message (p-value of 0.38). This suggests that expectations play a role in moral suasion and that preference effects cannot

---

<sup>12</sup>The unit of observation is the average contribution by group and message and we use a Wilcoxon signed-rank test for matched pairs given the lack of independence in behavior of subjects seeing different messages within the same 8 person group.

explain the whole effect of moral messages.

## 5.2 Experiment 4: Is there a preference effect?

In this section we study whether moral suasion has an effect on behavior that operates through preferences. In this experiment we hold fixed the message seen by a player's opponent, and compare the player's 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 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 expectations.

### 5.2.1 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 complicate inference about effects over preferences).<sup>13</sup> Finally, we adjusted the exchange rate to 8 EPs per dollar given the reduction in the number of rounds, so as to keep total average earnings at levels similar to those in experiment 1.

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

---

<sup>13</sup>Under 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.

Blank message).

### 5.2.2 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 learned 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).<sup>14</sup> This suggests that moral suasion affects behavior not only by affecting expectations but also by affecting 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, our results also indicate that the potential for persistent positive effects depends on the richness of the strategic environment in which moral suasion is used. In our experiment, the interaction of moral suasion and the presence of punishments appears important to sustain cooperation when moral messages or punishments alone could not do so.

---

<sup>14</sup>In 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.

An important additional question pertains to the mechanisms through which moral suasion operates. Our design allowed us to identify that moral suasion affects preferences. But moral suasion also seems to depend on whether players are confident that others have been “treated” as well, highlighting a social dimension of moral suasion linked to 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 social preferences such as those based on reciprocity motives is by now well known. However, the fact that social preferences can be leveraged to affect behavior through relatively cheap methods such as ethical discourse is intriguing, especially when considering that the provision of material incentives is costly. Future work should explore in more detail the variety of settings in which moral suasion 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.
- Andreoni, J. and J. Rao (2009), “The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism,” working paper University of California, San Diego.

- 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.
- Bohnet, I. and R.D. Cooter (2005). "Expressive Law: Framing or Equilibrium Selection?," working paper Harvard University."
- 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.
- Cooter, R. (1998). "Expressive Law and Economics," *Journal of Legal Studies* 27(2), 585-608.
- 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.
- Galbiati, R. and P. Vertova (2008), “Obligations and Cooperative Behavior in Public Good Games,” *Games and Economic Behavior* 64, 146-170.
- Geanakoplos, J., D. Pearce and E. Stacchetti (1989). “Psychological Games and Sequential Rationality,” *Games and Economic Behavior* 1(2), 60–79.
- 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	Suggestion	Golden Rule	Utilitarian		
Blank	0.161	0.455	0.053	0.007		
Nash		0.117	0.008	0.000		
Suggestion			0.019	0.004		
Round 11 versus Round 10						
	Nash	Suggestion	Golden Rule	Utilitarian		
Blank	0.052	0.009	0.000	0.015		
Nash		0.001	0.000	0.006		
Suggestion			0.020	0.208		

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. Tests are one-tailed. The alternative hypothesis is that the column treatment yields higher (lower, for the Nash column) contributions

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
	0.001	0.0002	0.0004	0.9652

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	0.636	0.9301
Golden Rule	0.012	0.5942

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	0.010

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 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

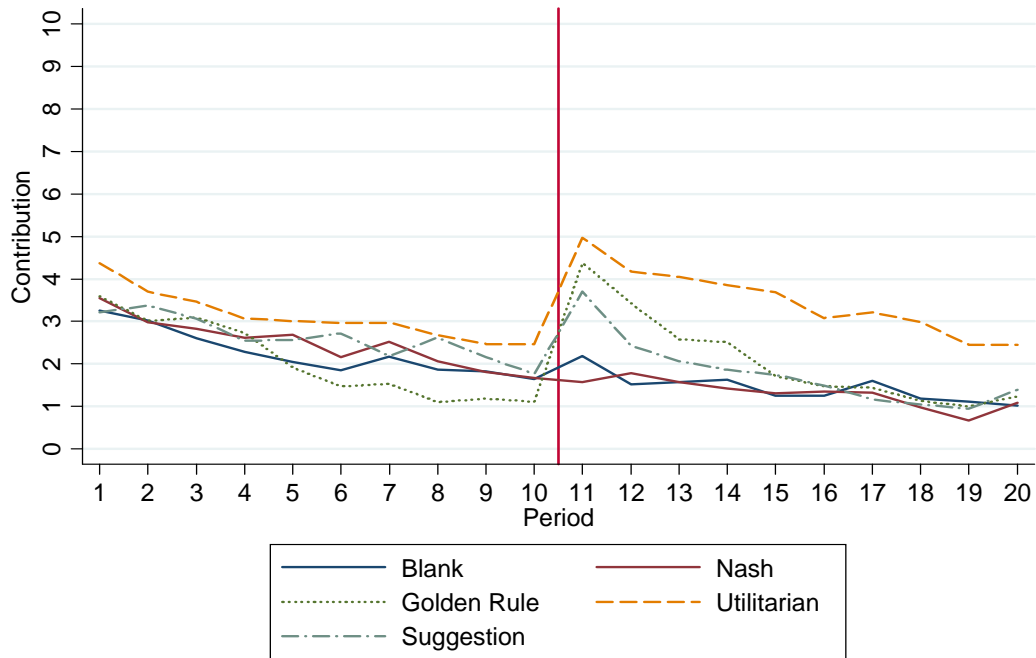
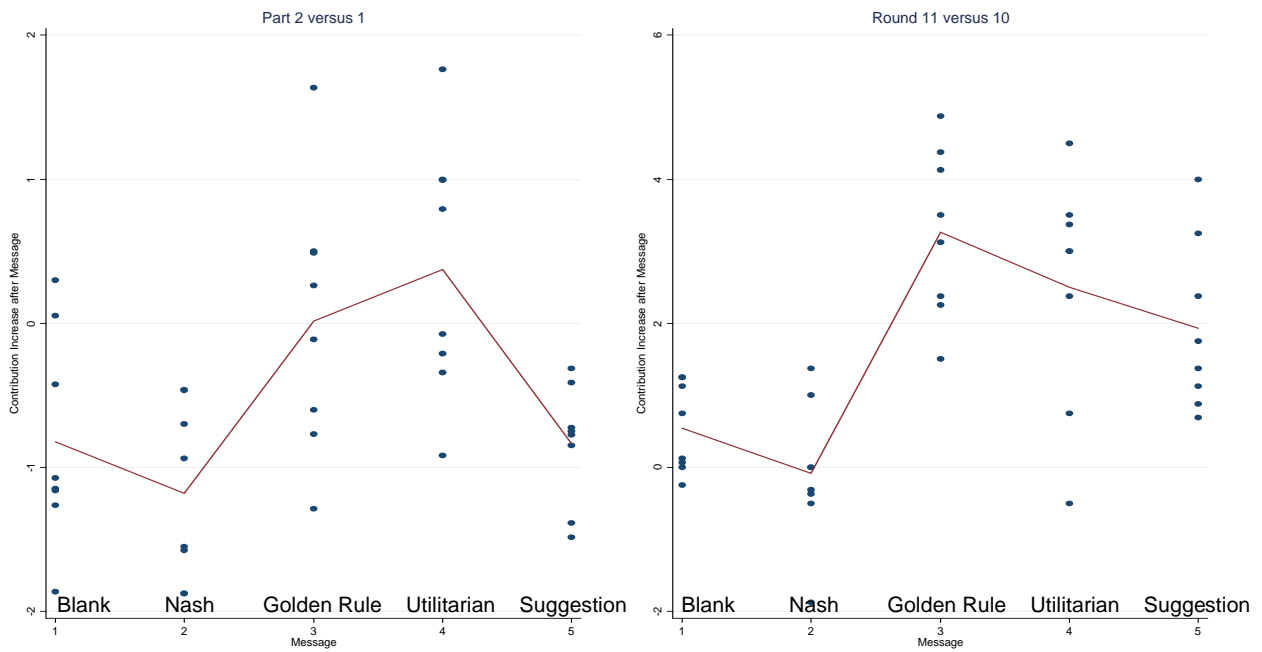


Figure 2: Change in Contributions after Message – Experiment 1



Note: Markers denote the change in average contributions by group and lines denote the change in average contribution by treatment.

Figure 3: Contributions by Round and Message – Experiment 2

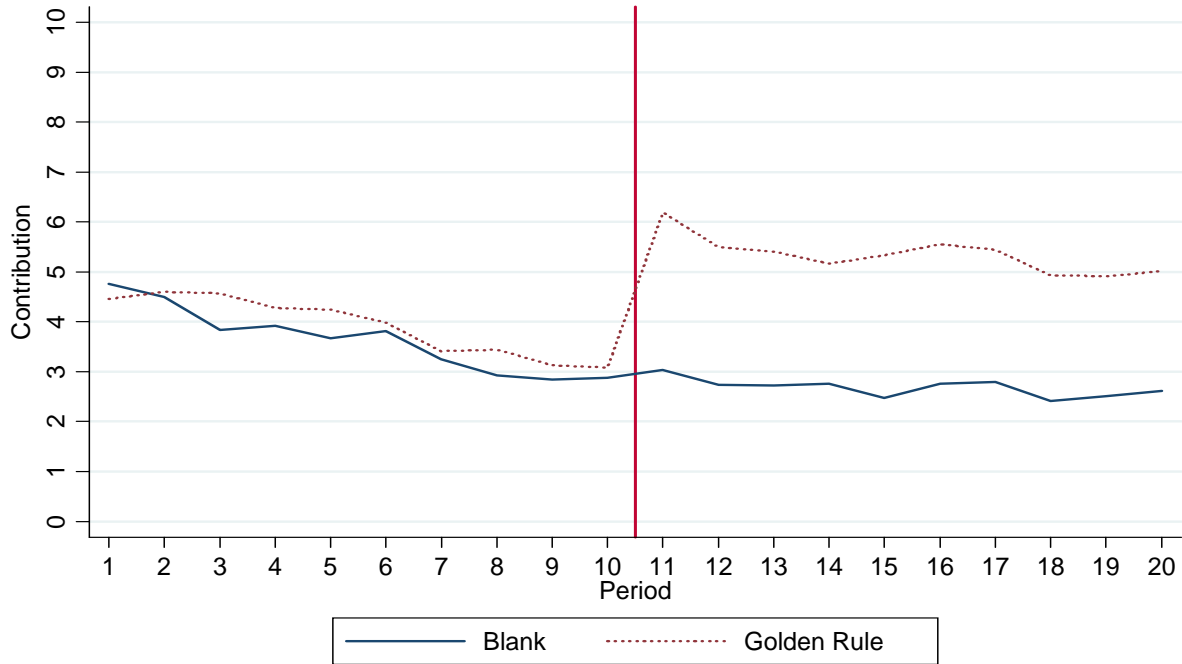


Figure 4: Change in Contributions after Message – Experiment 2

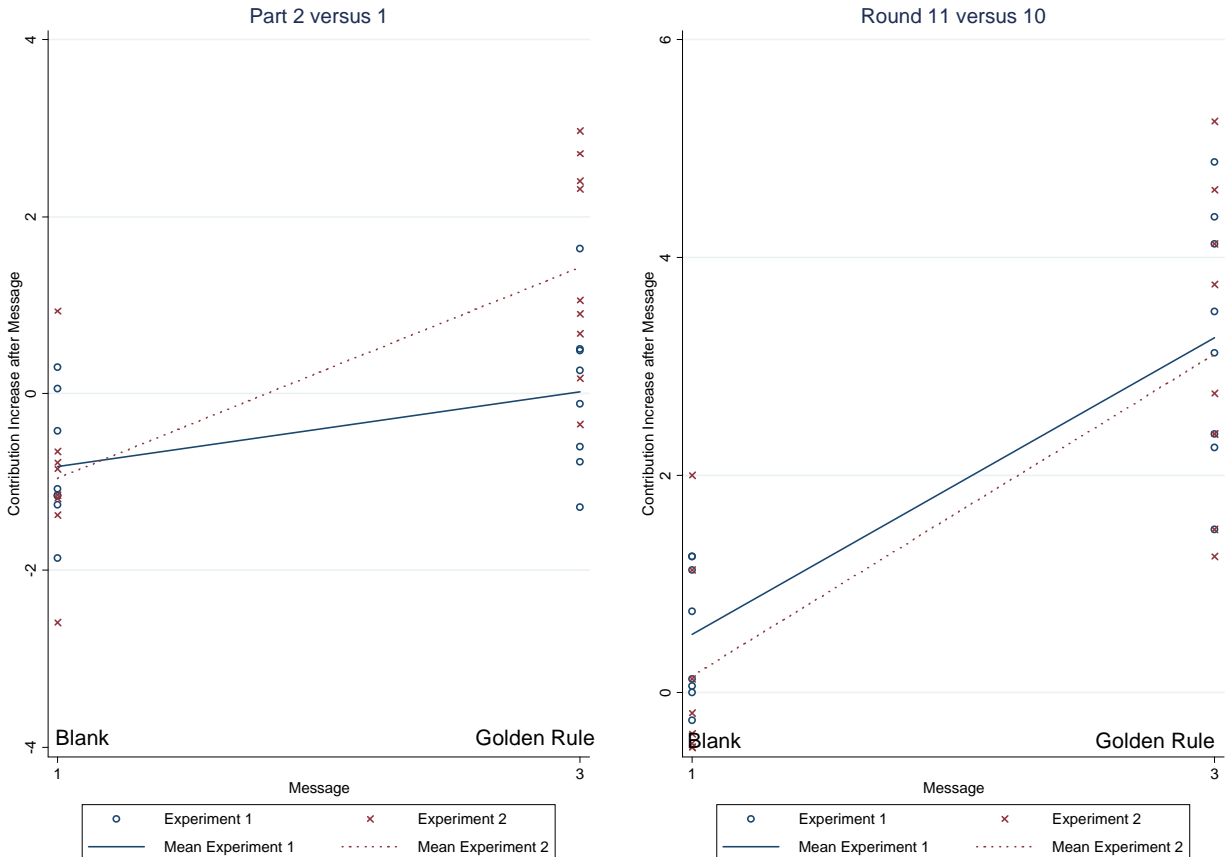


Figure 5: Contributions by Round and Message – Experiment 3

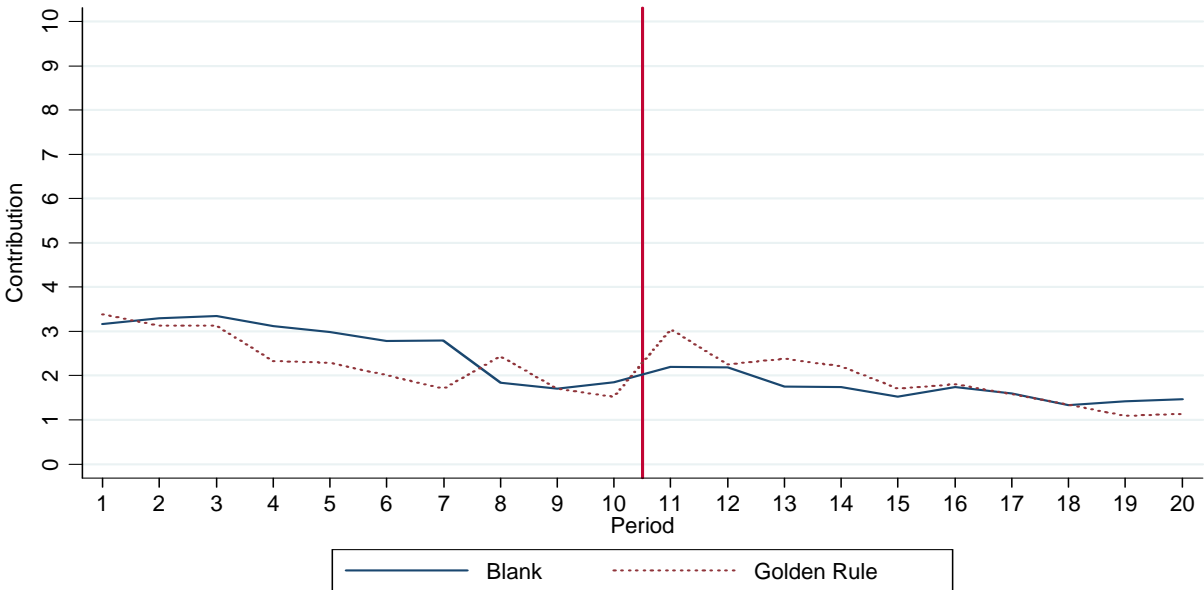


Figure 6: Change in Contributions after Message – Experiment 3

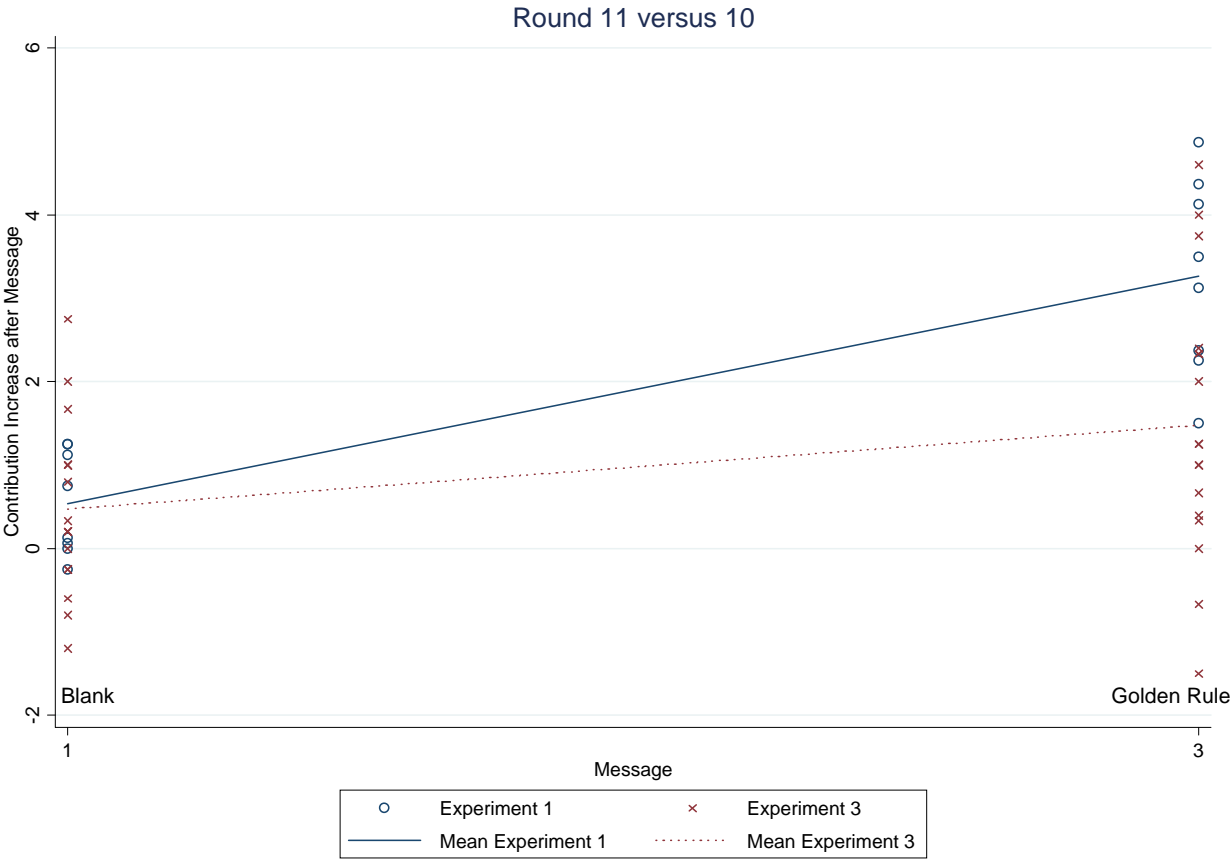


Figure 7: Contributions by Round and Message – Experiment 4

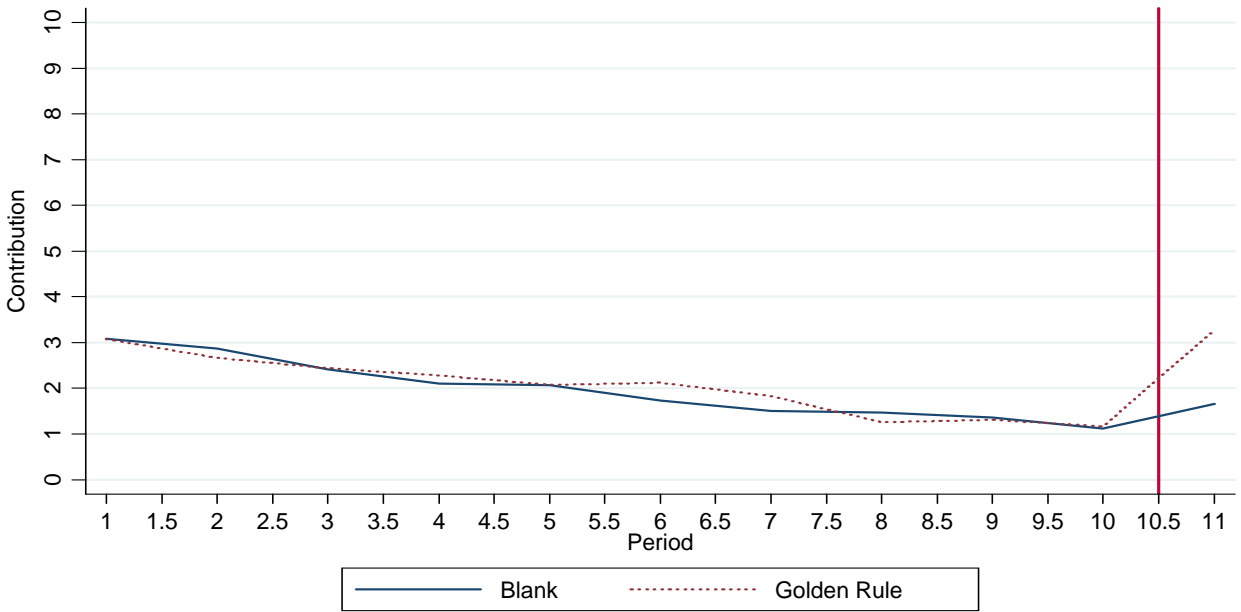
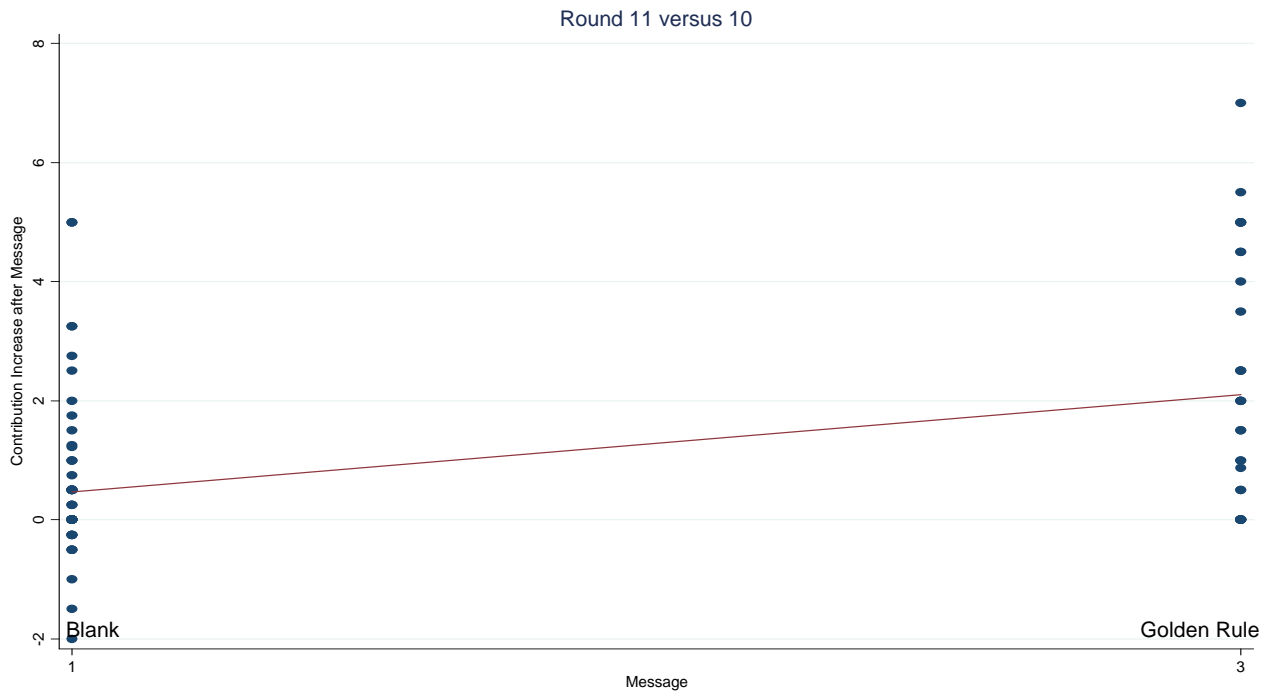


Figure 8: Change in Contributions after Message when Knowing that Other Saw Blank Message – Experiment 4



# Appendix: Experiment 1 Instructions

## (Not intended for publication)

### Welcome to the Experiment

---

This is an experiment on decision-making. How much you will be paid will depend on what you and others do during the experiment. Each of you will be paid at the end of the experiment by XLab personnel. All interactions in the experiment will be anonymous and through computer. Your identity will be kept totally private. Neither the experimenter nor the other participants will ever be able to tie you to your decisions.

---

#### Earnings

On top of the 5 dollars paid for participating, during the experiment you will earn “Experimental Points” (EPs) depending both on your actions and those of the other participants. At the end of the experiment the EPs you have earned will be exchanged into dollars at the rate of 1 dollar per 12 EPs.

#### The experiment

Participants will be randomly divided into groups of size 8. That means you will be randomly assigned to one group with seven other participants for the whole experiment. You will not know who these participants are, nor will they know who you are. You will then participate in 20 consecutive decision-making interactions. For each of the 20 interactions you will be randomly matched with one other participant from your group. After each interaction, you will be randomly matched again. After ten interactions, a screen will appear on your monitor containing a message. That message will be randomly selected by the computer from a set of messages including a blank message. At that point, please read the message carefully. The other members of your group will see the same message you see. After the message, the experiment will continue with another ten interactions. After the decision-making part of the experiment is over, you will complete a short questionnaire on the computer.

#### Each interaction

Each of the 20 decision-making interactions you will participate in will have the same structure. In each interaction you will receive an endowment of 10 EPs. A computer screen will give you the option to allocate your EPs between a personal account and a joint account with the other participant in that interaction. The EPs you allocate to your personal account go directly into your earnings. Every EPs allocated to the joint account by any of the two participants in an interaction yields 0.7 EPs to both participants. Thus, the earnings you obtain from the joint account depend both on the amounts that you and the other participant have allocated to it. Therefore, your individual earnings from each interaction can be calculated as follows:

$$\text{Earnings} = (\text{amount you allocate to your personal account}) + \\ + (0.7)(\text{amount allocated to joint account by you and the other participant}).$$

The following table displays the total earnings you obtain in one interaction depending on how many EPs you and the other participant in the interaction allocate to the joint account.

Your earnings in EPs		EPs allocated to joint account by the other participant										
		0	1	2	3	4	5	6	7	8	9	10
EPs allocated to joint account by you	0	10.0	10.7	11.4	12.1	12.8	13.5	14.2	14.9	15.6	16.3	17.0
	1	9.7	10.4	11.1	11.8	12.5	13.2	13.9	14.6	15.3	16.0	16.7
	2	9.4	10.1	10.8	11.5	12.2	12.9	13.6	14.3	15.0	15.7	16.4
	3	9.1	9.8	10.5	11.2	11.9	12.6	13.3	14.0	14.7	15.4	16.1
	4	8.8	9.5	10.2	10.9	11.6	12.3	13.0	13.7	14.4	15.1	15.8
	5	8.5	9.2	9.9	10.6	11.3	12.0	12.7	13.4	14.1	14.8	15.5
	6	8.2	8.9	9.6	10.3	11.0	11.7	12.4	13.1	13.8	14.5	15.2
	7	7.9	8.6	9.3	10.0	10.7	11.4	12.1	12.8	13.5	14.2	14.9
	8	7.6	8.3	9.0	9.7	10.4	11.1	11.8	12.5	13.2	13.9	14.6
	9	7.3	8.0	8.7	9.4	10.1	10.8	11.5	12.2	12.9	13.6	14.3
10	7.0	7.7	8.4	9.1	9.8	10.5	11.2	11.9	12.6	13.3	14.0	

After each interaction, the computer will show the allocations made to the joint account by you and the other participant, your earnings from that interaction, and your accumulated earnings. You will be able to advance from one interaction to the next by clicking the “continue” button at the bottom of your screen. Your earnings from all 20 interactions you participate in will be added up and will be exchanged for dollars at the end of the experiment.

## Overview of the experiment

- Participants will be randomly divided in groups of 8 for the whole experiment.
- You will participate in a total of 20 interactions.
- For each interaction, you will be randomly matched with one of 7 other participants in your group.
- When you start a new interaction, you will get randomly matched again.
- After 10 interactions, your computer screen will display a message randomly selected by the computer from a set of messages that includes a blank message. The 7 other members of you group will see the same message. You will participate in another 10 interactions after that.
- You will be paid your total earnings from all 20 interactions.
- 12 EPs are exchanged for 1 dollar.

## Conclusion

It is important that you understand the game and procedures of the experiment. Once the experiment begins, there will be no further discussion, and no communication of any kind among the participants is permitted, apart from the transmission of the decisions by computer. So please review the experiment in your mind now and raise your hand if you have any questions. Once all questions have been answered, the experiment will begin.

Now please double click on the Zleaf icon in your desktop, which should look like this:



Zleaf.exe

Thanks for participating.  
Good luck!