"Do the Right Thing:" The Effects of Moral Suasion on Cooperation

Ernesto Dal Bó — UC Berkeley & NBER

Pedro Dal Bó — Brown University & NBER*

May 2, 2014

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 also investigate the mechanisms through which moral suasion operates and find it affects both expectations and preferences.

JEL codes: C9, H41.

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

*For useful comments and suggestions we thank Alberto Alesina, Rachel Croson, Erik Eyster, Botond Koszegi, John Morgan, Santiago Oliveros, Parag Pathak, Louis Putterman, Matthew Rabin, Steve Tadelis, Dustin Tingley, 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 the right thing to do"), 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 focussed on measuring how material incentives can be manipulated to affect behavior. We have no equivalent body of knowledge on the effectiveness of moral suasion. In this paper we report on a series of experiments designed to measure 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 good game. We focus on this game because it neatly captures the clash between individual and collective rationality that motivates moral thinking and much of politics and public finance. In fact, normative political theory often presupposes that normative discourse can affect behavior, and align private and social objectives.[1] However, whether normative discourse can affect behavior, especially in strategic situations that capture the tension between the individual and the collective, is to the best of our knowledge unknown.

Establishing a causal role for normative statements faces important identification hurdles. The explanation of some historical event that invokes the prevalence of normative ideas meets

[1] For example, the influential notion of deliberative democracy has been described as hinging on a role for discourse about the common good to shape the attitudes of individual participants (see *inter alia*, Elster 1998, Ch. 1).

with the suspicion of endogeneity. Further investigation may always unveil deeper causes linked to the distribution of power, economic forces, or institutional factors, and suggest that the prevalence of normative ideas is itself the result of those deeper forces. These difficulties recommend the use of a laboratory study where exposure to normative ideas can be varied exogenously.

In this paper we adopt the experimental methodology to evaluate the effect of moral suasion on contribution levels in a public good game. We explore not only whether moral appeals affect behavior but also how moral suasion depends on aspects of strategic interaction, in an attempt to disentangle the mechanisms by which moral suasion operates.

In our first experiment we focus on establishing the existence of a moral suasion effect. Each session consisted of twenty rounds of a two-person public goods 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 for selfish preferences 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, which

provided a basic 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.[2] 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.

One reason why the moral suasion effect appears transitory is that in our first experiment participants could only punish low contributors by lowering their own contributions. To see if moral suasion effects can be persistent, in our second experiment we 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 (where punishment was not allowed), 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

[2] An experimenter's demand effect arises when subjects' behavior is affected by what they think the experimeter desires - see for example Zizzo (2010) for a discussion.

persistent cooperation, the interactive effect of punishments and a moral appeal did sustain cooperation at fairly high levels. This finding that moral suasion enhances the power of punishments is, to the best of our knowledge, novel.

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 mechanisms through which moral suasion operates.

A first possible mechanism through which moral suasion may operate is by affecting, perhaps temporarily, subjects' preferences. A moral message may raise the level of contribution 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. When a player does not care only about her own payoff, her beliefs about the contributions of others may affect her own contribution. Thus, it is possible that a moral message that is commonly observed may raise optimism about the contributions of others, and thus affect behavior.[3] 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

[3] As is well known, different forms of other-regarding preferences make beliefs about the behavior of others relevant to one's own contribution. On preferences that differ from material payoffs see Andreoni (1990), Rabin (1993), Fehr et al. (1997), Levine (1998) and Charness and Rabin (2002) among others. It is also well known that optimistic beliefs about the cooperation of others raise cooperation (Fischbacher et al. 2001).

indicates that the expectation effect is one way in which moral appeals work. It also 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. This result has an important implication for organizing the delivery of normative content to groups: it pays to ensure all members know everyone is receiving the same content.[4]

In order to determine whether a preference effect also operates, we conducted another experiment were 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.

The next section reviews related literature. Section 3 presents our first experiment and its results isolating the presence of moral suasion. Section 4 presents our results on the interaction between moral suasion and punishment, while Section 5 presents our third and fourth experiments investigating mechanisms. We conclude in Section 6 with a discussion of the limitations of our experiments, which can hopefully highlight valuable avenues for future research.

2 Related literature

The experimental study of cooperation boasts a rich tradition in economics and in political science. Most of the attention has been given to the power of extrinsic incentives to avoid the breakdown of cooperation. One example is the role of individual punishments in static

[4] Chwe (2001) tackles the issue of common knowledge in culture. He theorizes that the physical structure of many social interactions including rituals is optimized to ensure common knowledge, as a way to solve coordination problems. A complementary possibility is that such common knowledge amplifies the power of normative appeals.

environments (Ostrom et al 1992, Fehr and Gächter 2000). Another example is the role of reputational concerns in repeated interactions (Roth and Murnighan 1978, Dal Bó 2005). While there is much research on the power of extrinsic incentives to resolve cooperation problems, we have no equivalent knowledge on the role of moral suasion.

Some links have been established in political science between the spheres of the normative and the positive in relation to cooperation, with the connection going from the positive to the normative. For example, positive explanations for the development of norms have been given based on evolutionary considerations (e.g., Axelrod 1986).[5] The possibility that the normative may affect the positive is less well explored. And when normative notions have been investigated, the focus has been on the effect of existing preferences or beliefs on, for instance, selected criteria of justice (see Frohlich and Oppenheimer 1990, and references therein). What has been missing is evidence on whether exposure to moral ideas can *change* preferences or beliefs, and alter outcomes. The focus on change is important because it speaks to the potentialities of education and deliberation in promoting cooperation.

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). As Dawes and Thaler (1988) emphasize, it has been conjectured that one channel for the power of communication could be to highlight moral dimensions of behavior. However, research has shown communication to mainly act by creating mutual promises that emphasize an in-group vs out-group perspective (see Dawes and Thaler 1988 for a discussion and further references.) An advantage of our design is we can control subjects' exposure to distinctly ethical content. While our paper shows that communication with moral content from the experimenter can affect contributions, future research should study whether

[5] Axelrod conjectured that normative standards could affect behavior, however. In his influential book on cooperation, Axelrod (1984) considers several potential ways to promote cooperation, and devotes a subchapter (7.3) to the possibility of teaching people to care about the welfare of others. The extent to which this possibility is viable remains to be determined, and our study is a step in that direction.

communication with moral content among subjects has similar effects.

To our knowledge this paper is the first to report a laboratory study of the effect of moral suasion 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 the first to include moral suasion in experiments, however. 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.

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. 2013).[6] The exception is Schwarz and Orleans (1967), but their design confounds normative appeals with other factors that can affect compliance. It is worth noting that 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, which we take up in this paper.

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).[7] Interestingly, Dale and Morgan

[6] See also Apesteguia, Funk and Iriberri (2011) for a study of late returns in public libraries. They found that an email reminder has as much of an effect as a reminder which also mentions that timely returns matter to the functioning of the library system.

[7] On the effect of recommendations on coordination games see Van Huyck et al. (1992) and Brandts and

(2004) found that recommending the top contribution worked less well than recommending 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.

There is also a literature on how laws can express expected rules of behavior and may thus affect behavior –see Cooter (1995), Bohnet and Cooter (2005), Funk (2007), Galbiati and Vertova (2008, 2010) and Croson (2009). The closest paper in this literature is Galbiati and Vertova (2010) which shows that expressing an expected rule affects behavior both through changing beliefs about others and changing preferences. The power of "expressive law" may involve moral suasion among other forces, but it remains to be shown that moral suasion *per se* can have an effect. Our results that explicit moral suasion has effects reinforces the notion that expressive law could work through implicit moral suasion.

The results of our paper can be interpreted as capturing the effect of moral framing; see Andreoni (1995) on the effect of framing in public good games.[8]

3 Experiment 1: Does moral suasion affect cooperation?

This section covers an experiment that shows that exposure to moral appeals affects cooperative behavior.

MacLeod (1995).

[8]Framing plays a role in dictator games, too. Brañas Garza (2006) shows an increase in giving in dictator 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).

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 using the z-Tree software (Fischbacher 2007). 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. In each round, each player was randomly matched by the server to another person in the group. 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 on their computer screens, randomly selected by the server from a set of five possible messages. All subjects in the same eight-person 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 "Self-regarding"), 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 (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"). While the Blank and Suggestion messages were included as controls, the Self-regarding message was included to test whether the type of knowledge conveyed in a traditional economics class, in which self-regarding preferences are usually assumed, could lead to lower contributions.

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 use 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.[9] 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. Both moral messages as well as the morality-free suggestion to contribute included an added sentence stating "If you were to act accordingly, you would allocate 10 to the joint account." This wording surely has both pros and cons. The main disadvantage is

[9] 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 principles of the golden rule type), although it is an instance of a universalizable maxim.

that, if effects are found, one will wonder whether the results would persist with messages that do not spell out the full implications of the normative standard conveyed. The key advantage is that it ensures understanding by subjects, and we judged this to be the stronger consideration when taking a first step in this research agenda. 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. The objective of this paper is to investigate if at least one version of moral suasion can be effective, and not whether any or all moral communications will be. Future work should address variations in message wording.

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). The instructions for our experiments are available in an online appendix.

3.2 Results

The 320 subjects were divided in eight groups of eight people per message. 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.

To statistically compare changes in behavior throughout the paper we aggregate individual contributions at the level of the group and perform nonparametric tests. More specifically, we perform Wilcoxon rank-sum or sign-rank tests, as appropriate. We calculate randomized inference p-values, following Fisher (1935). We compute the pertinent Wilcoxon statistic,

and then obtain the p-values by re-assigning treatment status 10,000 times and deriving the distribution of the statistic. The p-value is the proportion of realized values of the statistic under treatment re-assignment that lie above (or below, when appropriate) the empirical value. We then double the p-values to make our tests two-tailed.

Figure 1 shows 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.[10] There are no significant differences in behavior across groups that ended seeing different messages, consistent with the random assignment of messages (the lowest p-value when comparing contributions in the pre-message across any two message conditions is .65).

In the context of this first experiment, we hypothesized that moral messages would increase contributions relative to the pre-message phase, and more importantly that it would increase contributions relative to all other messages. We hypothesized that the Suggestion message would increase contributions relative to the Blank and Self-regarding messages, and that the Self-regarding message would decrease contributions relative to the Blank message.

Did messages affect behavior as hypothesized? The answer is largely yes. We first consider the short run effects (round 10 vs round 11), and later discuss effects comparing behavior across the two sets of ten rounds before and after the message.

If we consider changes in contributions between rounds 10 and 11, Figure 1 shows that the Utilitarian message results in an increase of contributions from 2.47 to 4.97. This difference of 2.5 EPs, an increase in contributions of slightly over a 100%, is statistically significant with p-value of .015 - see panel A in Table 2. The Golden Rule message causes contributions to almost triple from round 10 to 11: as shown in Figure 1 contributions go from 1.11 to

[10] For a summary of the literature on public good games see Ledyard (1995) and Chaudhuri (2011).

4.38. This difference of 3.27 EPs is highly significant with p-value $< .01$. These changes in contributions are associated with respective increases in payoffs of 9% and 13% under the Utilitarian and Golden Rule messages (p-values of less than .01 and of .015 respectively). These effects, though perhaps not very large, are not trivial, especially when considering that moral suasion, unlike pecuniary incentives, is not costly in itself. But one can also see that a restart effect causes contributions to increase by more than half an experimental point from rounds 10 to 11 under the Blank message (from 1.64 to 2.18, with p-value of .07), while the Suggestion message increases contributions by almost two points (from 1.77 to 3.70, with p-value $< .01$). These increases under Blank and Suggestion correspond to 33% and 109% of the respective contribution levels in period 10. This means any analysis of the short run effect of moral messages must be compared with the effects of Blank and Suggestion messages to account for restart and demand effects. The p-values corresponding to the Wilcoxon tests for contribution differences across conditions in our first experiment are reported in panel B of Table 2.[11]

We find that both moral messages–Utilitarian and Golden Rule–generate increases in contributions from round 10 to 11 that are significantly higher than those under the Blank message (p-values of $< .001$ and .027, respectively –see Table 2, Panel B). In addition, 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 .031).[12]

To consider more than the short-run effects, we compare the average contributions in the entire second part of the experiment relative to the entire first, pre-message part. The

[11] The results are robust to performing statistical tests at the individual subject level clustering by group.

[12] Since 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. 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 Self-regarding message by contributing less than liberal subjects (p-value of .05 for average post-message vs pre-message contribution).

within treatment analysis in panel A of Table 2 reveals that the two moral messages are the only ones that do not display a statistically significant decrease in contributions in part 2. The analysis across treatments in panel B shows that the increase in contributions in Part 2 under the Utilitarian message is greater than the increase under the Blank message (p-values of .019) while the increase under the Golden Rule message is marginally insignificant under our two-tailed tests (p-value of .103; this increase would be significant with p-value .052 under a one-tailed test corresponding to our hypothesized direction of change). On the other hand there are no significant differences in terms of a decrease of contributions under Self-regarding or an increase under Suggestion relative to the Blank message (p-values of .34 and .89 respectively). Importantly, the increase in contributions under the moral messages is greater than under Suggestion (p-values of .036 and .007 for Golden Rule and Utilitarian respectively). This comparison has two implications. First, it is not just the recommendation of a given contribution level that affects behavior, but the explicitly moral part of the statement has an effect. Second, to the extent that the Suggestion message triggers similar experimenter demand effects, the overall effect of the moral messages cannot be attributed exclusively to an experimenter demand effect.[13]

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 .45 and .42 respectively). We found no long run effect of the Self-regarding message relative to Blank (p-value .34, see bottom of panel B in Table 2).

The Self-regarding message has a tiny and insignificant negative effect on contributions

[13] 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. All results in the paper are maintained if we focus on the percentage of subjects making the maximum contribution. The results are also maintained if we focus on subjects that remembered the message they saw.

from round 10 to 11 (p-value of .68, see panel A) and a negative short-run change relative to Blank that is marginally significant (p-value .09, see panel B of Table 2). One interpretation is that if the standard economics class-room content plays a role, it does not go beyond canceling out the restart effect under Blank. It should be noted that this effect may be constrained given that contributions are already quite low by round 10.

Our experimental instructions were clear that messages were randomly selected by the computer, in order to isolate an effect that is uncontaminated by inferences on the evolution of play that subjects could derive from observing a moral message. For example, if messages were not random, high contributors could interpret the moral message as a signal that contributions are lower than expected by the experimenter, infer that they are themselves contributing to much and respond by contributing less. Compatible with our design, we found no evidence that high contributors (those contributing more than the median in period 1) responded differently to the Utilitarian or Golden Rule messages (p-values of .9 and .13 respectively).

An interesting question 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; we leave this issue to future research. In the remainder of this paper we explore two other issues: the persistence of moral suasion and the mechanisms driving it.

4 Experiment 2: Moral suasion, punishment, and persistence

The main take away from our first experiment is that moral appeals can be used to affect cooperation, but that the effects of moral appeals appeared transitory. One interpretation is that moral discourse can be an effective, though short-lived, instrument to promote coop-

eration (incidentally, this might be a reason why exposure to normative messages in real life often takes on repetitive forms–e.g., attending mass once a week). Another interpretation is that players, though in principle willing to cooperate more in a persistent manner, eventually start to defect when they observe that not all players abide by the same principles. A player who wants to express frustration after witnessing bad behavior but also wishes to be cooperative has a single instrument, his contribution level, to pursue two different objectives. The retraction of cooperative behavior may be less common when subjects can use a separate instrument to punish players that have been uncooperative. Therefore, it is of interest to study moral suasion in a richer strategic environment, to see whether moral suasion triggers more persistent effects on 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.[14]

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. We focused on the Golden Rule message as opposed to the Utilitarian one because the Golden Rule condition exhibited the least persistence in our first experiment. Therefore, the Golden Rule message should provide a more stringent test for whether the addition of punishments affects the persistence of moral suasion. In addition, the Golden Rule is arguably a more universal message.

Second, the stage game was modified to allow subjects to punish their partner after seeing

[14]The availability of punishment may have stronger effects on contributions when there are more than the two players we consider here, as changing contribution levels affects all players but punishments can be targeted to a particular player.

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 six experimental sessions at XLAB, UC Berkeley with a total of 136 subjects. Eight groups of eight people saw the Blank message and nine groups saw the Golden Rule message. 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.

Figure 2 shows the evolution of contributions to the joint account by round and message. As before, 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-values of .55 for rounds 1 to 10 and .87 for round 10).[15] Contribution levels before subjects see the messages are greater than in experiment 1, when punishments were not available. This difference is significant (p-value: $< .01$). However, it is interesting to note that these contributions decrease with experience. In fact, the contributions in round 10 are significantly smaller than in round 1 (p-value $< .01$). In other words, while punishments help raise contributions in the absence of moral messages, they cannot prevent a significant erosion of cooperation.

Did messages affect contributions in the presence of punishment? We begin by examining

[15] Surprisingly this is not always the case for punishments. 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: .24) but it is significant in round 10 (p-value of .01). Given the controlled nature of the experiment we attribute this imbalance to a random occurrence.

the short run effects. As shown in Figure 2, moral suasion has a large effect on contributions. While the average contribution under the Blank message increased by 0.15 EPs from round 10 to 11 (around 5% and statistically insignificant with p-value of .82 –see panel A of Table 3) the average contribution under the Golden Rule condition increased by 3.11 EPs (roughly a 100% increase and highly significant with p-value $< .01$). Moroever, the change in contributions between rounds 10 and 11 under the Golden Rule message was significantly higher than under the Blank message (p-value of .001 –see panel B of Table 3 for the Wilcoxon rank-sum test results across treatments).

More importantly, the moral suasion effect is persistent in this second experiment. The average contribution suffers a decrease from round 10 to round 20 under the Blank condition of 0.26 EPs (near a 10% increase and p-value of .45, see panel A of Table 3), while the average contribution under the Golden Rule condition *increases* by 1.93 EPs (a 62% increase and p-value of 0.015). From Figure 2 and Table 3 we see that, considering 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 in the impact of the messages is significant (p-value $< .01$ for all rounds, see panel B of Table 3).

The persistence of moral suasion in the second experiment is also reflected in panel C of Table 3. The increase in contributions caused by the moral message from Part 1 to Part 2 is significantly larger in this experiment than in the first one (p-value of .019). That panel also shows that adding punishments did not change the effect of the Blank message (p-values of .14 for rounds 10 and 11 and .78 for all rounds). 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 .87). This means that the impact of allowing punishments is not on the short run effect of moral suasion, but on its persistence. 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 –see Figures 1 and 2. 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 (see the fourth column in panel B; p-value of $< .01$). However, if we focus on rounds 10 and 11 (third column in panel B) we find the opposite.[16] 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 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.[17]

[16] Consistently with the previous literature, we find that subjects tend to punish subjects that contributed less but there is also perverse punishment (by subjects who punish partners that contributed more). See Fehr and Gächter (2000), Anderson and Putterman (2006) and Carpenter (2007).

[17] 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 the behavior of 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.[18] 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 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.

[18] Moral suasion may also change expectations about others' expectations about one's own behavior. The latter case involves higher order beliefs that 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.

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. We chose the Golden Rule message to maintain consistency with both the first and second experiments.[19] 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 could not be higher than 50%. This was a true statement, since messages were randomized within group with equal probability. 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 six experimental sessions at XLAB, UC Berkeley with a total of 136 subjects. There were 17 groups of eight subjects; 69 subjects saw the Blank message and 67 saw the Golden Rule message. 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.

Figure 3 shows 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 differ-

[19] Future research should examine potential differences between the Golden Rule and Utilitarian messages in terms of the relevance of the expectations effect; the frames these messages introduce could draw attention to the actions of others to different degrees.

ences in behavior across subjects that ended seeing different messages, consistent with the random assignment of messages.

From Figure 3 and panel A of Table 4, 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 only significant for the Golden Rule message (p-value of .002) and it is significantly larger than that under Blank (p-value for the differential effect under Golden Rule is .018 as reflected in Panel B of Table 4).[20]

Did messages affect behavior differently than in our first experiment? To answer this question we focus only on round 11. In that round, subjects in the Golden Rule condition in this experiment only differ from those in the first experiment in terms of their confidence that their partner has seen the same message. After round 11 there could be an additional effect present: subjects in this third experiment also face a different experience of play (they may encounter players that have seen a different message and, hence, behave differently). To eliminate this second effect we restrict attention to round 11.

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 .01, see panel C in Table 4), while there are no differences for the Blank message (p-value of .76). This suggests that expectations play a role in moral suasion and that preference effects cannot 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 op-

[20] 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.

ponent, 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 (we favored the Golden Rule over the Utilitarian message again to maintain consistency with all previous experiments). Second, the choice of messages and matching of subjects within a group of eight was such that half the subjects could be truthfully informed that their opponent had seen the Blank message. Two of these four "informed" subjects saw the Blank message and two saw the Golden Rule message, and all of them were paired with subjects exposed to a Blank message. Subjects knew that if they were informed of the 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).[21] 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.

[21] 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.

To test whether moral suasion has an effect through preferences, we compare the behavior of those who received a Blank message against those that received the Golden Rule message, while restricting attention to "informed" subjects so as to hold constant the subjects' information about the (Blank) message seen by their partners.[22]

5.2.2 Results

We conducted ten experimental sessions at XLAB, UC Berkeley with a total of 264 subjects. Of these, 132 subjects saw the Blank message and received no information about the message seen by their partner. The other 132 were informed that their partner had seen the Blank message and saw themselves a Blank message in 66 cases, and the Golden Rule message in the other 66 cases. 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.

Figure 4 shows 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 panel A of Table 5 we see that both the Blank and Golden Rule messages result in an increase in average contributions from round 10 to 11 (there is again a small restart effect). However, this increase is only significant for the Golden Rule message (p-value below .001),

[22] An alternative method to identify pure preference effects could be to use the strategy method, as in Fischbacher et al (2001). However, if higher order beliefs matter, following the strategy method we could mistakenly attribute to preferences what is in fact the result of higher order beliefs.

and it is significantly greater than the increase under the Blank message (the p-value for the differential effect of the Golden Rule message is below .001 as reflected in panel B of Table 5).[23] 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 organizations and politics.

Our results 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.

An important additional question pertains to the mechanisms through which moral suasion operates. Our design allowed us to identify that moral suasion is stronger when players are confident that others have been "treated" as well, highlighting a social amplification of moral suasion linked to expectations about others. When preferences are either purely pecuniary or based on a strictly individual moral imperative those expectation-driven effects should not arise. Their emergence suggests that moral suasion leverages a pro-social, but also reciprocity-based, aspect of preferences. Our experiments show that moral suasion operates also by affecting preferences, holding expectations constant. In other words, moral suasion affects what players expect from each other, but also "who they are" in terms of preferences.

[23] 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 Wilcoxon sign-rank test for matched pairs.

The existence of social preferences such as those related to 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–involving laws and regulations, monitoring, and money–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.

Some limitations of our experiments highlight yet other avenues for future research. Our experiments involved a restricted, and especially worded, set of messages, and therefore our results do not imply that every form of moral suasion will succeed. Thus, it would be important to investigate the power of messages when the full implications of the normative standard are not spelled out, as well as situations where messages nominally invoke morality without actually conveying a moral standard or justification. It is possible that moral suasion affects both the levels of contributions players consider appropriate (a moral imperative), but also the utility value for meeting that imperative. Distinguishing which of these two elements is affected by moral discourse remains an open question, as is whether it is expectations at the behavior or the expectations of others that matters. Lastly, it would be important to investigate moral suasion in settings other than public good games, both in the lab and the field.

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.

Axelrod, R. (1984). The Evolution of Cooperation, New York: Basic Books.

Axelrod, R. (1986). "An Evolutionary Approach to Norms," *American Political Science Review* 80(4), 1095-1111.

Apesteguía, J., P. Funk and N. Iriberri (2011). "Promoting Rule Compliance in Daily Life: Evidence from a Randomized Field Experiment in the Public Libraries of Barcelona," working paper Universitat Pompeu Fabra.

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.

Brañas-Garza, P. (2006). "Promoting Helping Behavior with Framing in Dictator Games," *Journal of Economic Psychology* 28(4), 477-486.

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.

Chaudhuri, A. (2011). "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics* 14(1), 47-83.

Chwe, M. S-Y. (2001). "Rational Ritual. Culture, Coordination and Common Knowledge. Princeton: Princeton University Press.

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.

Croson, R. (2009). "Experimental Law and Economics," *Annual Review of Law and Social Science* 5, 25-44.

Dale, D.J. and J. Morgan (2004). "Fairness Equilibria and the Private Provision of Public Goods," mimeo UC Berkeley.

Dawes, R. and R. Thaler (1988). "Anomalies: Cooperation," *The Journal of Economic Perspectives* 2(3), 187-197.

Elster, J. (1998). Deliberative Democracy, Cambridge University Press.

Fehr, E. and S. Gächter (2000). "Cooperation and Punishment in Public Goods Experiments," *American Economic Review* 90(4), 980-94.

Fehr, E., S. Gächter and G. Kirchsteiger (1997). "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica* 65(4), 833-860.

Fellner, G., R. Sausgruber, and C. Traxler (2013). "Legal Threat, Moral Appeal and Social Information: Testing Enforcement Strategies in the Field." Forthcoming, *Journal of the European Economic Association.*

Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics* 10(2), 171-178.

Fischbacher, U., S. Gächter and E. Fehr (2001). "Are People Conditionally Cooperative? Evidence From a Public Goods Experiment," *Economics Letters* 71, 397–404.

Fisher, R. A. (1935). The Design of Experiments, London: Oliver and Boyd.

Frohlich, N. and J. Oppenheimer (1990). "Choosing Justice in Experimental Democracies with Production," *American Political Science Review* 84(2), 461–477.

Funk, P. (2007). "Is There An Expresive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines," *American Law and Economic Review* 9, 135–159.

Galbiati, R. and P. Vertova (2008), "Obligations and Cooperative Behavior in Public Good Games," *Games and Economic Behavior* 64, 146-170.

Galbiati, R. and P. Vertova (2010), "How Laws Affect Behavior: Obligations, incentives and cooperative behaviour," forthcoming, *International Review of Law and Economics.*

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.

Rabin, M. (1993). "Incorporating fairness into game theory and economics. American Economic Review 83(5), 1281–1302

Roth, A. and J.K. Murnighan (1978). "Equilibrium behavior and repeated play of the prisoner's dilemma," *Journal of Mathematical Psychology* 17(2), 189-198.

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.

Zizzo, D.J. (2010). "Experimenter demand effects in economic experiments," *Experimental Economics* 13(1), 75-98.

Table 1: Treatment Messages

	Name	Message
1	Blank	Blank
2	Self-regarding	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: Changes in contributions by treatment

	Message				
	Blank	Self-regarding	Golden Rule	Utilitarian	Suggestion
Round	(1)	(2)	(3)	(4)	(5)
Round 11 - Round 10	0.54	-0.09	3.27	2.50	1.93
P-value	0.066	0.679	0.008	0.015	0.006
Part 2 (post-message) - Part 1 (pre-message)	-0.82	-1.08	0.01	0.37	-0.84
P-value	0.040	0.008	0.941	0.384	0.006
Number of subjects	64	64	64	64	64

Panel B: Differences in changes in contributions across treatments

Round 11 versus Round 10 - P-values

	Self-regarding	Suggestion	Golden Rule	Utilitarian
Blank	0.092	0.010	0.000	0.027
Self-regarding		0.002	0.000	0.010
Suggestion			0.031	0.361

Part 2 versus Part 1 - P-values

	Self-regarding	Suggestion	Golden Rule	Utilitarian
Blank	0.344	0.890	0.103	0.019
Self-regarding		0.219	0.012	0.001
Suggestion			0.036	0.007

Note: P-values in Panel A rely on non-parametric matched pairs tests. The null hypothesis is that contributions before and after the message within treatment stem from the same distribution. The matched pairs involve contribution levels before and after the message, taking the average contribution of each 8-person group on each side of the message as a single observation. In Panel B we test the null hypothesis that the change in contributions from part 1 to part 2 or from round 10 to 11 for groups in a row treatment category stem from the same distribution as the changes for groups in the column treatment category, relying on a non-parametric Ranksum test. We treat 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: Changes in contributions and punishments by treatment

	Contributions		Punishments	
	Blank	Golden Rule	Blank	Golden Rule
	(1)	(2)	(3)	(4)
Round 11 - Round 10	0.15	3.11	0.47	-0.29
P-value	0.820	0.004	0.153	0.342
Round 20 - Round 10	-0.26	1.93	-0.30	0.06
P-value	0.446	0.015	0.283	0.976
Part 2 - Part 1	-0.83	1.43	-0.03	0.86
P-value	0.058	0.013	0.849	0.004
Number of subjects	64	72	64	72

Panel B: Differences in changes in contributions and punishments across treatments

	Contributions		Punishments	
	Round 10 vs 11	Part 1 vs 2	Round 10 vs 11	Part 1 vs 2
P-value				
Blank-Golden Rule	0.001	0.001	0.079	0.001

Panel C: Comparison between experiments 1 and 2 - Are changes in contributions different?

	Round 10 vs. 11	Part 1 vs. 2
P-values Blank - Exp 1 vs 2	0.140	0.783
P-values Golden Rule - Exp 1 vs 2	0.865	0.019

Note: P-values in Panel A rely on non-parametric matched pairs tests. The null hypothesis is that contributions before and after the message within treatment stem from the same distribution. The matched pairs involve contribution levels before and after the message, taking the average contribution of each 8-person group on each side of the message as a single observation. In Panel B we test the null hypothesis that the change in contributions or punishment from part 1 to part 2 or from round 10 to 11 for groups in different treatment categories stem from the same distribution. In Panel C we test the null that the change in contributions from part 1 to part 2 or from round 10 to 11 in each treatment stem from the same distribution across experiments 1 and 2. In panels B and C we treat the change in the average contribution of each 8-person group as a single observation, and perform non-parametric Ranksum tests.

Table 4: Do expectations play a role? - Experiment 3

Panel A: Changes in contributions by treatment

	Message	
	Blank	Golden Rule
	(1)	(2)
Round 11 - Round 10	0.35	1.52
P-value	0.118	0.002
Part 2 - Part 1	-0.99	-0.51
P-value	0.004	0.076
Number of subjects	69	67

Panel B: Differences in changes in contributions across treatments

	Round 10 vs 11
P-value	
Blank-Golden Rule	0.018

Panel C: Comparison between experiments 1 and 3 - Are changes in contributions different?

	Round 10 vs. 11
P-values Blank - Exp 1 vs 3	0.761
P-values Golden Rule - Exp 1 vs 3	0.010

Note: P-values in panels A and B rely on a non-parametric matched pairs test. In Panel A we test the null that contributions before and after the message within treatment stem from the same distribution. The matched pair involves contributions before and after the message, where we take the average contribution at the group-message level as a single observation. In Panel B we test the null that the change in contributions from round 10 to 11 in both treatments stem from the same distribution, and the matched pair is changes in contributions from round 10 to 11 by subjects within a group, with each pair element corresponding to different message treatments. Panel C performs nonparametric Ranksum tests on the null that the changes in contributions under each treatment stem from the same distribution across experiments. We treat the change in the average contribution at the group-message level as a single observation.

Table 5: Are There Preference Effects? - Experiment 4

Panel A: Changes in contributions by treatment

	Message	
	Blank	Golden Rule
	(1)	(2)
Round 11 - Round 10	0.26	2.10
P-value	0.670	0.000
Number of subjects	66	66

Panel B: Differences in changes in contributions across treatments

	Round 10 vs 11
P-value	
Blank-Golden Rule	0.000

Note: Observations are restricted to subjects who knew their partner had seen the Blank message. P-values in both panels correspond to non-parametric matched pairs tests. In Panel A we test the null that contributions before and after the message within treatment stem from the same distribution, and the matched pair is contributions before and after the message taking the average contribution at the group-message level as a single observation. In Panel B we test the null that the difference across treatments in the change in contributions from round 10 to 11 stems from the same distribution, and the matched pair is changes in contributions from round 10 to 11 by subjects within a group, with each pair element corresponding to different message treatments. Again we treat the average contribution at the group-message level as a single observation.

Figure 1: Contributions by Round and Message – Experiment 1

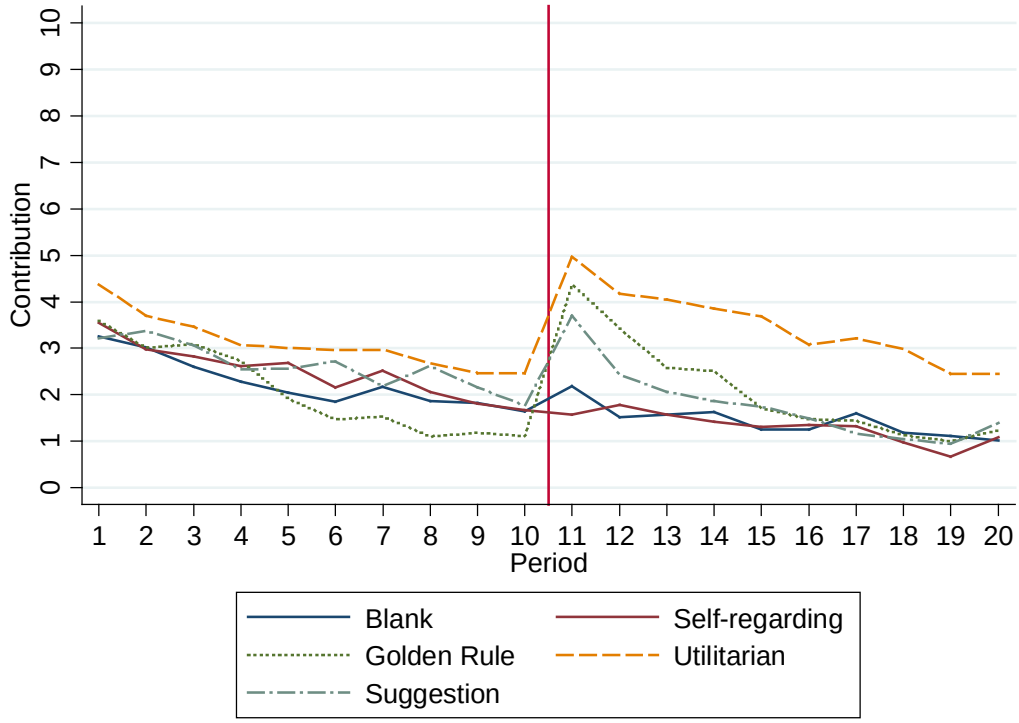

Figure 2: Contributions by Round and Message – Experiment 2

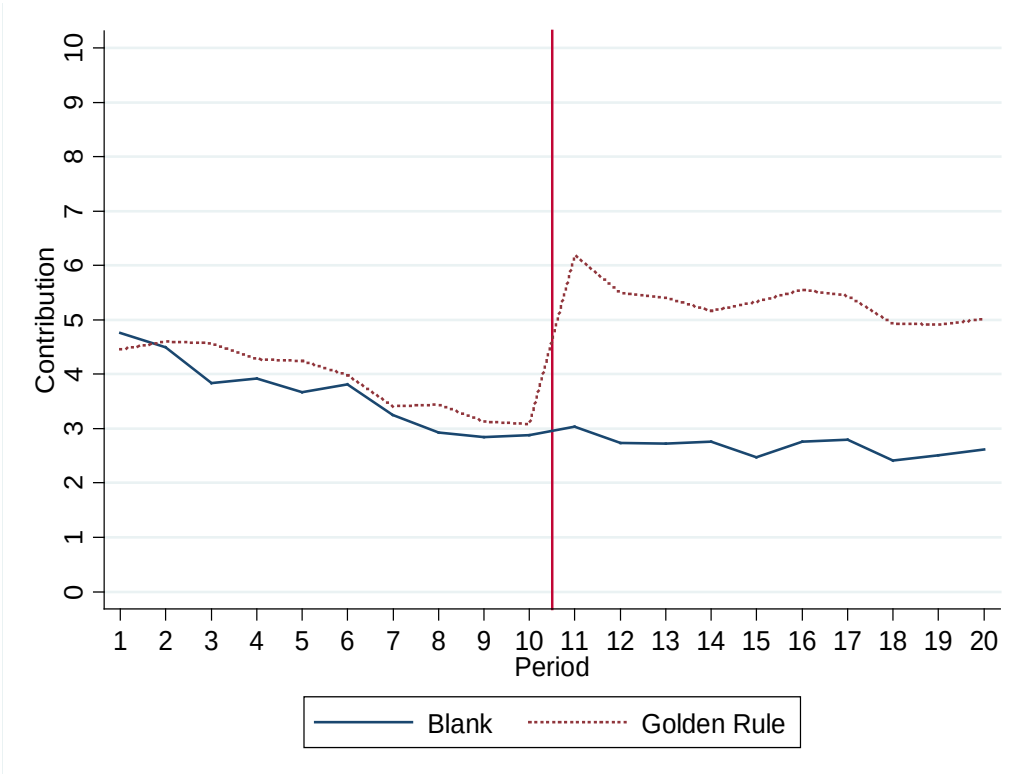

Figure 3: Contributions by Round and Message – Experiment 3

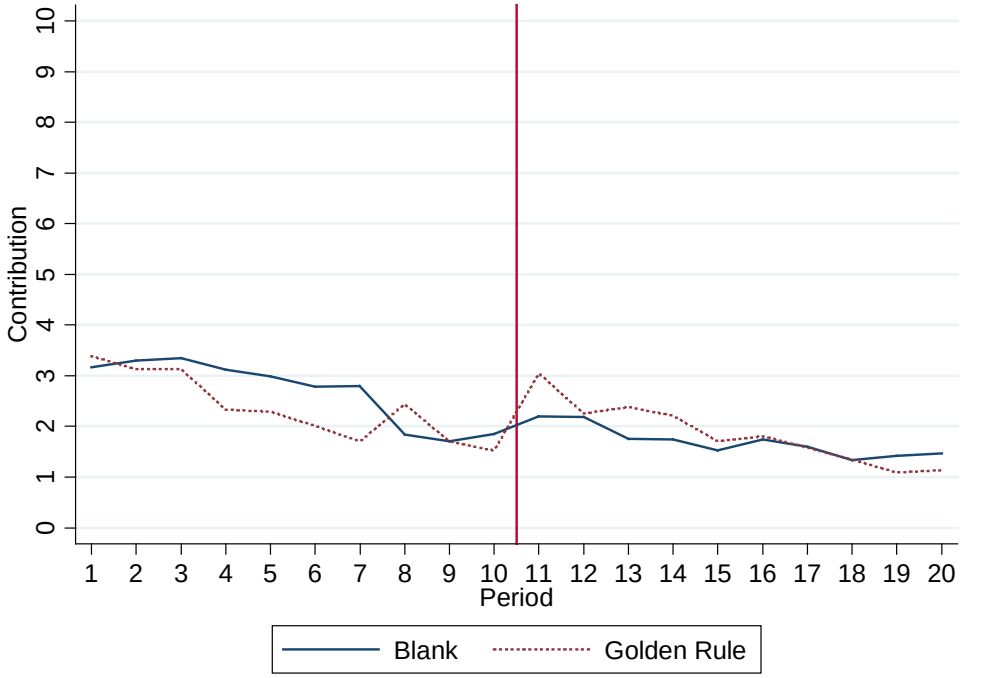

Figure 4: Contributions by Round and Message – Experiment 4
(Subjects who know partner saw Blank)