Just Ask:
Preference Revelation and Lying in a Public Goods Experiment

Andrea Robbett*

July 8, 2016

Abstract

This paper studies the extent to which simply asking group members to report their demand-type promotes cooperation in a public goods experiment with private, unobservable incentives. The design uses a simple public goods game, in which group members differ in both their dominant strategy contributions and socially optimal contributions. Treatments are conducted in which individuals either do not report their private payoff information, report only to group members, report to group members who can punish them, or report to a binding mechanism that charges them the socially optimal contribution for their message. In all cases, messages are non-verifiable and participants are told that they are free to lie. When participants are able to report their type, either to their group members or to a binding mechanism, they contribute more. Further, the misreporting of type is less frequent, considered more dishonest, and punished more harshly than free-riding. Consistent with work showing that weak punishment can backfire, the punishment mechanism is underutilized in this environment and its presence negates the positive effect of sharing information.

JEL Classification: H41, C92, C72
Keywords: Experimental Economics, Public Good Provision, Collective Action, Cooperation, Preference Revelation

*Department of Economics, Middlebury College, Middlebury, VT 05753. arobbett@middlebury.edu. (802) 443-5653. Fax: (802) 443-2080. Many thanks to Jeff Carpenter, Peter Matthews, conference participants at the 6th Biennial Social Dilemmas Conference at Brown University and ESA Heidelberg, as well as seminar participants at Maastricht University, the University of Amsterdam, Erasmus University, Middlebury College, Bates College, Union College, University of Oslo, University of Cologne, Max Planck Institute for Research on Collective Goods, and LMU Munich. Funding provided by Middlebury College. The author reports no conflict of interest. The experiment reported in this paper was approved by the Middlebury College Institutional Review Board.

1 Introduction

Can simply asking individuals to reveal their willingness to pay for a public good lead to higher provision? Traditional public economic theory suggests that it will not, as agents will strategically misreport how much they value the public good. The question of how to provide public goods efficiently is a central focus of public economics, largely due to this difficulty in observing individual preferences. Experimental public goods games, on the other hand, have largely focused on how to promote voluntary contributions when incentives are *common knowledge*, and have not yet directly addressed whether free-riding in this type of game is analogous to the strategic under-reporting of demand. This paper seeks to unite these two literatures, by introducing a simple, novel public goods game in which participants are assigned different payoff types and given the opportunity to share their type, either with group members or with a central "mechanism" that naively implements the optimal outcome based on what it is told. The results indicate that *asking participants to report their demand can significantly increase contributions* and that the decision to *under-report demand* is *not* the same as the decision to *free-ride by under-contributing*. These findings suggest that surveying households about their demand can generate significantly higher provision than relying only on voluntary contributions.

In theory, the efficient allocation of public goods is impeded by two closely-related free-rider problems. First, agents who are asked to voluntarily contribute have incentive to disregard the full benefits to the group and to under-contribute relative to the social optimum. The question of how to promote voluntary contributions has been a primary focus of public goods experiments. If agents' incentives are known, such that the efficient allocation can be computed, then this problem can be overcome by taxes that implement the optimal provision or by institutions that promote prosocial behavior, such as peer enforcement (e.g. Fehr and Gächter 2000). The second impediment is the information, or preference-revelation, problem, which occurs when individual utility functions are *not* observable and has been a primary focus of public economics and mechanism design. Without knowledge of individual utility functions, it may be impossible for a central tax authority to calculate and implement the optimal allocation or for peers to recognize and enforce cooperative behavior. In this case, agents who are

asked to reveal or signal their demand, for the purposes of computing costs, have incentive to misrepresent their preferences and under-report how much they truly value the public good. These two types of free-riding behavior – under-contributing and under-reporting – are typically taken to be analogous and the assumption that agents will misreport their demand according to their narrow self-interest is generally taken as the starting point for the design of mechanisms to incentivize truthful revelation (Green and Laffont 1977; or see Krajbich, Camerer, Ledyard, and Rangel 2009 for a recent discussion).[1]

The question remains, however, as to the extent to which individuals *will* strategically misrepresent their demand in a public goods game when directly asked. Will individuals reveal their preferences truthfully, even in the absence of a mechanism incentivizing them to do so? Is the decision to free-ride by *under-reporting* one's demand equivalent to the decision to free-ride by *under-contributing*? Does the ability to share private information enable groups to enforce higher contributions? Previous experimental work provides reason to believe that people might be hesitant to misreport, even in situations when they would otherwise choose to free-ride. In particular, an abundance of recent experimental evidence suggests that individuals may have an inherent aversion to lying and, across a wide variety of contexts, often truthfully share private information against their monetary interests (Fischbacher and Föllmi-Heusi 2013; Gneezy 2005; a review follows in the next section). If individuals are averse to misreporting their type, either due to psychological lying costs or strategic considerations, then the link between "demand revelation by report" and "demand revelation by contribution" could be broken and this finding would suggest that simply asking agents to share private information can go some way toward solving the preference revelation problem without appealing to more complicated mechanisms.

The experiment reported in this paper tests whether, and the circumstances under which, asking participants to reveal their demand can lead to more efficient public good provision in a

[1]This link goes back to Samuelson (1954)'s treatment of public expenditures, which demonstrates that decentralized market systems cannot efficiently allocate public goods. Samuelson points that out that individuals may be asked to reveal "preferences by signalling in response to price parameters or Lagrangian multipliers, to questionnaires, or to other devices. But ... any one person can hope to snatch some selfish benefit ..." and "now it is in the selfish interest of each person to give *false* signals, to pretend to have less interest in a collective consumption activity than he really has, etc." (pp. 388 - 389). Work on the design of mechanisms typically assumes that individuals will not take into account the "truthfulness" of their message, with some notable exceptions, such as Green and Laffont (1986) and Deneckere and Severinov (2008), which introduce limitations on the message space that can reflect lying costs or the extent to which agents can exaggerate without being caught.

social dilemma game with unobservable incentives. The design provides a direct test of whether participants are more willing to free-ride by voluntarily under-contributing than they are by purposefully misreporting their demand. The experiment further tests whether sharing private information with group members, both with and without punishment opportunities, leads to similar revelation levels and whether this information can enable groups to enforce higher levels of voluntary contributions.

Participants play a 10-period public goods game, in which group members may differ from each other in both their individually optimal and socially optimal contribution levels. Specifically, each participant is assigned to be either a "Blue Type" (an agent with high demand for the public good) or a "Red Type" (an agent with low demand). The existence of two types and the payoffs of each are common knowledge but individuals do not know the composition of their group or the types of specific group members. In the baseline version of the game (VCM), participants learn their type in each period and then choose how much to voluntarily contribute to the public good. The game was carefully designed to be as similar as possible to the standard, linear public goods game widely used in experimental work, with similarly transparent incentives for the participants.[2] The game described in this paper preserves each of the fundamental features of that game while introducing heterogeneity of incentives. First, each type faces a social dilemma, such that their individually-optimal contribution is less than their socially-optimal contribution. Second, both the individually-optimal and socially-optimal contributions are in dominant strategies, such that they do not depend on beliefs about the contributions of others. This was particularly important in this setting, in which beliefs about others' contributions would depend on beliefs about the composition of the group. Finally, the two types receive the same payoff as each other under the dominant strategy outcome and under the socially optimal outcome. Therefore, concerns about payoff inequality in mixed groups should not influence behavior or differentially push the outcome toward the Nash equilibrium

[2]In the typical public goods experiment, participants receive an endowment and can decide how much to keep for themselves (returning a payoff of 1) or to contribute to a public account (returning a payoff less than 1 on each unit contributed by a group member). The parameters are set such that a player whose objective is to maximize his own payoff would contribute nothing while a player whose objective is to maximize group payoffs would contribute everything. Further, these monetary incentives are in dominant strategies and thus do not depend on beliefs about the decisions of others. Typically participants in this game begin by contributing approximately half of their endowment on average, but contributions decline quickly over time unless supported by mechanisms such as punishment or communication (Ledyard 1995).

or toward the social optimum.

In addition to the baseline VCM condition, three other experimental conditions are conducted in which individuals have the ability to share information by sending an unverifiable message about their type. The first, *Mechanism*, directly tests whether participants are similarly willing to free-ride by under-reporting a type to a mechanism as they are by under-contributing in the standard voluntary contribution game. The Mechanism condition was designed to be strategically isomorphic to the baseline VCM. Rather than simply entering how many tokens they wish to contribute, participants in the Mechanism condition enter a message about their demand type (their color) and then are automatically charged the socially optimal contribution for someone of their reported type. To keep the strategy set identical across the two conditions, there exists a complete menu of messages that allows them to make any of the possible contributions available to VCM participants. Therefore, the only difference between the two conditions is that participants in the Mechanism condition who wish to free ride must enter a message that they know to be untrue.

The experiment further tests whether the ability to send messages about types can enable the group members themselves to overcome the information problem and enforce higher contributions endogenously. In the final two conditions, participants also report a type, but can then voluntarily contribute any amount. This message is reported only to their group members, who are shown each individual's message and contribution at the end of the period, and does not otherwise affect their payoffs.[3] The *Revelation* condition tests whether simply asking participants to share their type, even without any enforcement mechanism, is sufficient to increase contributions. The *Punishment* condition is identical to the Revelation condition, but adds a peer enforcement stage in which participants can pay to reduce the earnings of their group members after viewing each member's contribution and message. Thus, if participants use the punishment mechanism to enforce contributions that match the social optimum for the message, then the Punishment condition should work similarly to Mechanism – or better if participants are more willing to accurately report their types to their peers. Alternatively, if

[3]In all four conditions, the contribution of each group member is displayed at the end of each period. For the Mechanism condition, the contribution always directly corresponds to the message the individual sends. In the other two message conditions, the link between the individual's message and contribution is broken.

the punishment mechanism is weak, its presence could backfire. Previous literature has shown that sanctioning mechanisms are used less frequently when cooperation is observed with noise (Greiner and Ambrus 2012) or when payoffs are private information (Robbett 2014) and that weak punishments and fines can have a negative effect (Gneezy and Rustichini 2000; Fehr, Gächter, and Kirchsteiger 1997; Aquino, Gazzale, and Jacobson 2015).

Overall, the experiment provides evidence that simply *asking participants to report their type can increase cooperation* in public goods games with private information. The paper reports three main findings. First, despite being strategically equivalent, the Mechanism condition generates higher contributions than VCM. In other words, *the decision to under-report demand is not the same as the decision to free-ride*: Participants are more willing to under-contribute relative to the social optimum than they are to under-report their type to a mechanism that will charge them the optimal contribution for their report.

Second, participants reporting their type to their group members are significantly more likely to report the truth than those who are reporting to a binding mechanism. Participants in both the Revelation and Punishment conditions typically report their types truthfully – between 70% and 83% of the time. Finally, the truthful reports also translate into higher voluntary contributions. Participants in the Revelation condition, who have the opportunity to report their type to their group members, contribute significantly more than participants in the VCM baseline and similarly to participants in the Mechanism condition. However, the positive effect of reporting one's type is undermined in the presence of peer punishment. Contributions in the Punishment condition do not differ from the VCM baseline and are significantly lower than in Revelation, despite messages being equally truthful. The expected punishments are not sufficient to change the incentives of free-riders and, instead, are greatest for participants who send transparently false messages, further suggesting that dishonesty is viewed more harshly than free-riding.

The remainder of Section 1 reviews the related literature; Section 2 describes the payoffs and conditions; Section 3 presents the main findings; Section 4 provides a further discussion of the main results within the context of the previous literature and post-experiment questionnaire data, as well as discussing possible implications.

1.1 Related Experimental Results

Experimental work investigating solutions for efficient public good provision has followed two largely separate strands: the introduction of pro-social institutions, such as communication and punishment, that can promote public good provision by relying on instrinsic prosocial motivations of participants, and the design of mechanisms to extrinsically incentivize the truthful revelation of private information. As the experimental public goods literature is far too vast to review here, what follows focuses on these strands and their connection to the current investigation. To my knowledge, this experiment is the first to directly address whether, and the circumstances under which, asking participants to report their demand can reduce free-riding in a social dilemma with private incentives.

First, a large literature has investigated whether and when prosocial motivations of group members can be sufficient to overcome the individual monetary interest to free-ride. In groups with homogeneous financial incentives, access to costly punishment often substantially increases contributions. Participants in these games, who are able to monitor the contributions of their group members and pay to reduce the earnings of others, do use the mechanism to punish free-riders, and free-riders increase their contributions in response (Fehr and Gächter 2000). When contributions are observed with noise (e.g., Ambrus and Greiner 2012) or when a participant's ability to have contributed is unknown (e.g., Bornstein and Weisel 2010), punishment is used less frequently yet is generally still successful in increasing contributions. Pre-play communication among group members also consistently reduces free-riding – although participants are generally forbidden from sharing any private payoff information during these conversations (Issac and Walker 1988, Chan, Mestelman, Moir, and Muller 1999). Robbett (2014 working paper) assesses whether punishment or unrestricted communication can be similarly successful at promoting cooperation in public goods games when group members have heterogeneous monetary incentives, using a payoff structure similar to the baseline experiment in this paper. This preliminary evidence suggests that free-form communication *does* still promote cooperation in this game, especially when participants are permitted to refer to their types or payoffs in their discussions. However, punishment on its own *does not* increase contributions in groups with private incentives, where participants are unable to distinguish between cooperative versus

uncooperative behavior. Yet, if it is the case that participants tend to truthfully and believably report their types, then this uncertainty would be resolved and punishment could again provide an effective mechanism when combined with the ability to share information, as in the Punishment condition of the current paper. Alternatively, participants who can be held accountable by their group members may be less willing to share honestly.

The second approach is the design of mechanisms to incentivize truthful revelation of demand. Experimental tests of demand-revealing mechanisms have found that, even if participants have a dominant strategy to reveal their demand, they frequently do not play it and may instead converge toward other, weakly dominated, messages (Attiyeh, Franciosi, and Isaac 2000, Kawagoe and Mori 2001). However, participants do tend to follow best response dynamics (Healy 2006; Cason, Saijo, Sjöström, and Yamato 2006), which can lead to convergence in supermodular games (e.g., Chen and Plott 1996). Chen (2008) provides a comprehensive review of the experimental mechanism design literature, while Healy (2006) contributes a model of best response dynamics that successfully predicts equilibrium (non)convergence in an experimental test of five different mechanisms. Since these types of experiments are designed to test whether extrinsic incentives alone can generate truthful revelation, they typically abstract away from framing the situation as a social dilemma or from framing reports as true vs. false. In contrast, the current paper considers an environment in which participants are *not* incentivized to tell the truth, but in which both the truthfulness of different messages and the consequences of those messages for other participants are fully transparent. There are exceptions, however, either that involve a social dilemma interaction or that frame one possible message as truthful. For instance, Falkinger, Fehr, Gächter, and Winter-Ebmer test the Falkinger Mechanism, which subsidizes above average contributors and taxes below average contributors, in a standard voluntary contribution mechanism game, although the mechanism is based on contributions rather than reports. Rondeau, Schulze, and Poe (1999) ask participants to contribute to a provision point mechanism (in which a public good is provided only if an undisclosed contribution level is reached), with a money-back guarantee if the project is underfunded and a proportional-rebate if it is overfunded; they find that, on average, participants contribute more than their induced value. Krajbich, Camerer, Ledyard, and Rangel (2009) display a dollar value for the public

good to subjects in an fMRI scanner and ask them to report back whether they have a high or low value, with subjects paying a penalty if their response differs from a prediction based on their measured brain activity. Given the accuracy of the experimenters' predictions, subjects are incentivized to truthfully report their values and nearly always do.

The Mechanism condition reported in this paper shares commonalities with experimental work on tax compliance, in which participants are asked to report an income, to be taxed, with some chance of being audited and fined if the reported income is inaccurate. These experiments have found that participants are more likely to truthfully report their income than they are to make an analogous decision in a lottery choice with the same parameters (Baldry 1986), and that the inclusion of a public good, such that participants receive a return from their tax payments, increases compliance (Alm, Jackson, and McKee 1992).

Finally, as mentioned in the introduction, a wide variety of experimental work has shown that participants often truthfully report private information to group members or to the experimenter against their own material interest. For instance, experimental participants who must report to the experimenter either their performance or random outcomes that determine their pay typically do not maximally inflate their scores (Fischbacher and Föllmi-Heusi 2013; Cohn, Fehr, and Marechal 2014; Mazar, Amir, and Ariely 2008). Senders in sender-receiver games often report the true state of the world against their own self-interest (Gneezy 2005; Lundquist, Ellingsen, and Johannesson 2009), even when lying would be beneficial to both players (Erat and Gneezy 2012; Cappelen, Sørensen, and Tungodden 2013). Participants in bilateral bargaining games often reveal their private values and costs truthfully (Ellingsen, Johannesson, Lilja and Zetterqvist 2009) and participants who are permitted to discuss payoff information during free-form chat periods of public goods games frequently tell the truth (Robbett 2014).

2 Experimental Design

2.1 Payoffs in a Simple Public Goods Game with Heterogeneous Incentives

Participants played a 10-period public goods game in groups of three. As in the typical public goods experiment, participants simultaneously chose the amount that they wanted to contribute

to a group account, which benefited all members. Participants then received a payoff that depended on: the amount that they personally contributed; the amount contributed by their group members; and their payoff type – each individual was assigned to be either a Blue Type or a Red Type. Participants knew about the existence of the two types and received paper payoff tables for both types at the start of the experiment. Types were randomly assigned at the start of each period. They always knew their own type for the period, but not the types of their group members. In order to generate different dominant strategy contributions and socially optimal contributions for the two types, the payoffs must necessarily depart from the linear payoff function used in the standard game.[4] However, the payoffs otherwise preserve most of the features in the typical game.[5]

Own Contribution

Total Contributions of Other Two Group Members	0	1	2	3	4
0	20	25	30	25	20
1	30	35	40	35	30
2	40	45	50	45	40
3	50	55	60	55	50
4	60	65	70	65	60
5	70	75	80	75	70
6	80	85	90	85	80
7	90	95	100	95	90
8	100	105	110	105	100

Own Contribution

Total Contributions of Other Two Group Members	0	1	2	3	4
0	10	5	0	0	0
1	20	15	10	0	0
2	30	25	20	0	0
3	40	35	30	0	0
4	50	45	40	0	0
5	60	55	50	0	0
6	70	65	60	0	0
7	80	75	70	0	0
8	90	85	80	0	0

Figure 1: The Payoffs for a Blue Type (Left) and a Red Type (Right)

The payoffs for the two types are displayed in Figure 1 and will be described in detail below. The payoffs were constructed with several considerations in mind, all of which serve to align the present game with the typical set-up. First, it was important that the incentives for each type be as transparent as possible. Thus, rather than giving participants payoffs that

[4]The vast majority of public goods experiments employ a linear payoff function, developed by Isaac, Walker, and Thomas (1984). In this game, participants receive a payoff equal to the amount of their endowment that they kept for themselves plus the total amount contributed multiplied by factor less than one. The linear structure implies corner solutions for both the individual's dominant strategy contribution (0 if the multiplier is less than 1 and everything otherwise) and the socially optimal contribution (0 if the multiplier is less than 1/Group Size and everything otherwise). Thus this payoff structure cannot be used to create interior dominant strategies or interior optimal strategies.

[5]This payoff structure was previously used in one earlier experiment (Robbett 2014 working paper), which found that contributions in the baseline game with types that are fixed (not randomly re-generated as in this paper) exhibited similar patterns to the standard, linear public goods experiments: Contributions started at 40% of the social optimum and then declined significantly over time. This consistency provides a "sanity check" that participants interpret the incentives of this slightly more complicated game similarly to the standard linear public goods game.

required them to understand either complicated payoff functions or extensive payoff tables, the tables provided were as compact and transparent as possible. Only two types were used, such that participants could easily internalize and keep in mind the incentives for both type of player. Participants also played practice rounds in all-Red and all-Blue groups to learn about the incentives for each. Second, *both* types faced a social dilemma, such that the individually optimal contribution was less than the socially optimal contribution. Third, neither of these contribution levels – that is, the individually-optimal level or the socially optimal level – depends on the contributions of other members. This is a critical feature, as it implies that beliefs about the composition of the group or the cooperativeness of group members do not affect the individual's monetary incentives.

The dominant strategy for each type can be read directly off of the table. Looking across any row for the Blue Types (the left table in Figure 1), increasing own contribution from 0 to 1 or from 1 to 2 *increases* payoffs by 5 points each, while increasing contribution from 2 to 3 or from 3 to 4 *decreases* payoffs by 5 points. This is true in every row (i.e. for any level of contributions by one's group members). Thus Blue Types have a dominant strategy of contributing 2. Likewise, the Red Types' payoffs decrease by 5 points for each token they contribute up to 2 and then they receive 0 points for contributing beyond 2. Thus they have a dominant strategy of contributing 0.

Next, note that, for each extra token contributed by another group member, points increase by 10. This is true for both the Blue Types and the Red Types and can be seen by looking down any column in the two tables. Therefore the social benefit to contribution does not depend on beliefs about the composition of one's group: contributing an extra token always increases both other group members' payoffs by 10 points each. Therefore, a Blue Type who considers increasing his contribution beyond his dominant strategy of 2 would pay a cost of 5 points per token, but would benefit his group by a total of 20 points per token. Thus the social optimum requires that Blue Types contribute 4 and, by similar reasoning, Red Types contribute 2.

Finally, the payoffs were calibrated to be the same for the two types under the dominant strategy outcome and under the socially optimal outcome. For instance, imagine that there are

two Blue Types and a Red Type in the group. If all players play their dominant strategy, Blue Types receive a payoff of 50 (own contribution = 2 and total contribution of the two others = 2) and Red Types receive a payoff of 50 (own contribution = 0 and total contributions of the two others = 4). Alternatively, if all players play the socially optimal strategy, Blue Types receive a payoff of 80 (own contribution = 4 and the total contribution of the two others = 6) and Red Types receive a payoff of 80 (own contribution = 2 and total contributions of the two others = 8). The equality of payoffs across types under both the dominant strategy and socially optimal outcomes holds for every possible group composition. This keeps the payoffs in line with the standard, homogeneous game and eliminates concerns that inequality aversion might differentially push the group toward either outcome.

2.2 Experimental Procedure

Overall, 192 Middlebury College undergraduate students participated in the experiment. Four experimental conditions were conducted: VCM, Mechanism, Revelation, and Punishment. For each condition, three separate sessions were conducted and each condition contained between 15 and 18 independent groups of three. Participants interacted using the experimental software z-Tree (Fischbacher 2007) and had dividers separating their computer terminals.

In the baseline *Voluntary Contribution Mechanism (VCM)* condition, each participant simultaneously chose how many of four tokens to voluntarily contribute to a group account. At the end of each period, participants learned their payoffs and total contributions. They also saw three boxes on their screen, each associated with one of the group members and displaying that individual's contribution for the period. The order of the boxes was randomized each period, so that no individual's behavior could be tracked across periods.

The *Mechanism* condition was designed to be strategically isomorphic to the VCM condition and differed only in that participants were asked to make a contribution by reporting a type, rather than reporting a number. After being informed of their type for the period, participants saw a screen asking "What is your type?" and were required to write a message of the form "I am a [Color] Type," with several possible colors to choose from. No other messages were accepted and the participant could not proceed until a valid message was entered. If a par-

ticipant reported that he was a Blue Type, he was then automatically required to contribute 4 tokens, the socially optimal contribution for Blue Types. If a participant reported that she was a Red Type, she was then automatically required to contribute 2 tokens, the socially optimal contribution for Red Types. In order to keep the strategy set identical between the VCM and Mechanism conditions, participants also were given access to messages that mapped to the other possible levels of contributions: Participants who reported that they were "Green" contributed 0, "Yellow" contributed 1, and "Purple" contributed 3. This mapping between messages and contributions was fully transparent and always appeared on the participants' screens directly below the box into which they typed their messages. A screenshot of this stage is provided in Figure 2. Furthermore, it was emphasized in the instructions and throughout the experiment that participants were free to be untruthful in their messages, that all messages were well within the rules of the game, and that group members would only ever see the messages, never the true types.[6] Of course, if, for instance, a Red Type reports that he is "Green" then it would be clear to the group that he is not being truthful, as Green Types do not exist. However, this is identical to the situation in the VCM condition in which a Red Type would be revealed to be a free-rider if he contributed 0.

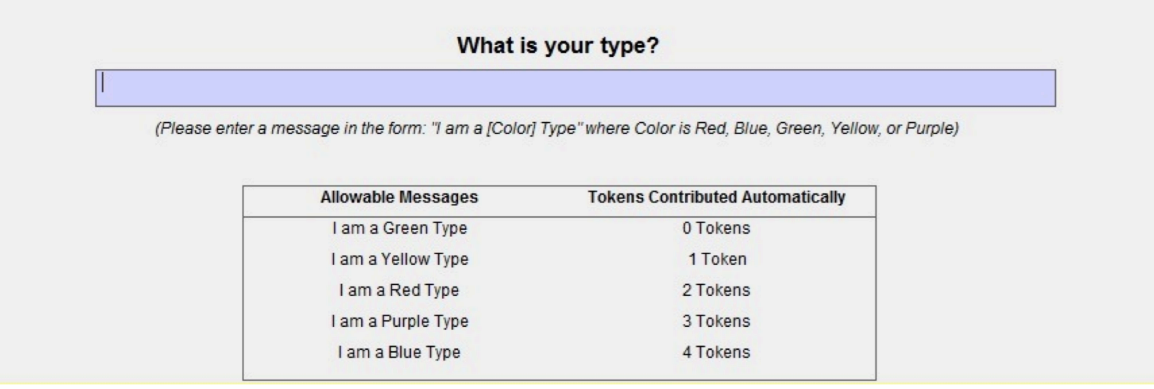

Figure 2: Screenshot of Message Screen in the Mechanism Condition

The *Revelation* condition was similar to the Mechanism condition, in that participants were again asked to report a type and an identical set of messages was available. In this condition, however, the link was broken between the message and the contribution: after reporting

[6]The full set of instructions for all conditions are included as an Appendix.

a type, the participant was free to voluntarily contribute any amount he or she wished. The same messages were available to the participants: Red and Blue, but also Green, Yellow, and Purple. In this, case, however, note that the latter three options are equally meaningless and thus reporting one of these types is equivalent to refusing to reveal information. The three boxes that the participants saw at the end of the period indicated both the message that the person reported as well as his/her contribution. Requiring participants to enter only one of the five available messages, rather than giving them access to an unrestricted chat phase, eliminates any pro-social effects of communicating with one's group members or discussing the game. Instead, this condition isolates the effect of sharing one's type with the group and allows for a direct comparison with the Mechanism condition. Likewise, displaying each group member's revealed type alongside each contribution at the end of the period, rather than prior to the contribution stage, aligns this condition with the Mechanism condition and eliminates the possibility that reports would be interpreted as "promises" or would otherwise influence the expectations of group members prior to contribution (see Charness and Dufwenberg 2006 and 2011).

The *Punishment* condition was identical to the Revelation condition, with the addition of a costly peer punishment phase. After participants viewed the message-contribution pair of each individual in the group, they could pay to reduce the earnings of one or both of their group members. They could pay as much as they liked to reduce the earnings of a group member and the earnings of that participant was then reduced by *three* times the amount paid, a common punishment ratio that is typically highly effective at promoting contributions (Nikiforakis and Normann 2008).

Prior to participating in the experiment, participants received the payoff tables and detailed explanations of how to read them (see Part 1 instructions in the Appendix). They were then quizzed on the tables and participated in a practice round of 5 periods in which all participants were Red Types and 5 periods in which all participants were Blue Types, which gave them the opportunity to learn about the game and the incentives. It was emphasized that the incentives for both types remained the same in mixed groups. All participants played the VCM version during the practice rounds, to ensure that this experience was identical across conditions. They were then matched into *new* groups for the experiment and received instructions for their

specific condition. Following the experiment, participants completed a brief questionnaire and were paid, in cash, the sum of their earnings over all ten periods at the rate of 50 points = 1 US dollar, in addition to a 5 dollar show up fee.

3 Results

The primary question is whether being asked to directly reveal one's demand, either to a binding mechanism or to one's group members, affects cooperation. This question is addressed by the regression models estimated in Table 1. Rather than taking the participants' token contribution as the dependent variable, we exploit the fact that the two types face a largely symmetric choice over different ranges of the contribution space. Specifically, both types have a dominant strategy contribution, but could contribute two more tokens to achieve the social optimum, which carries a marginal cost of 5 and marginal social benefit of 10. Therefore, the dependent variable used in Table 1 is the participant's contribution, beyond their payoff-maximizing contribution, as a percent of the distance to the social optimum, i.e., Percent Optimal Contribution = (Contribution - 2)/(4-2) for Blue Types and (Contribution - 0)/(2-0) for Red Types. This variable thus captures the extent to which the participant cooperates, beyond their self-interested contribution, and is equal to 1 when the participant cooperates fully and 0 when the participant free-rides fully.[7] Standard errors are clustered at the level of the participant's group, such that there are 64 clusters overall. The baseline VCM is the omitted condition.

First, we see that contributions are significantly higher in the Mechanism condition, in which participants contribute by reporting a type, than in the VCM condition, in which participants simply make a contribution. Specifically, asking participants to contribute by reporting their type to a mechanism, which charges them the optimal contribution for their report, increases contributions beyond the VCM baseline by nearly 15% of the distance from the Nash outcome to optimum. This corresponds to an increase of 73% and the difference is significant at the $p < .05$ level.

[7]The distribution of actual token contribution by type and condition is reported in the Appendix.

Table 1: Percent Optimal Contribution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mechanism	0.145**	0.136*	0.136*	0.272***	0.160**	0.150**	0.150**	0.311***
	(0.0716)	(0.0737)	(0.0735)	(0.0986)	(0.0699)	(0.0719)	(0.0719)	(0.0927)
Revelation	0.145*	0.164**	0.164**	0.162*	0.142*	0.158**	0.160**	0.188**
	(0.0791)	(0.0731)	(0.0730)	(0.0877)	(0.0767)	(0.0709)	(0.0709)	(0.0901)
Punishment	-0.0323	-0.0239	-0.0228	-0.0727	-0.0173	-0.0113	-0.0108	-0.0381
	(0.0559)	(0.0539)	(0.0539)	(0.0744)	(0.0515)	(0.0502)	(0.0504)	(0.0690)
Female		-0.0883	-0.0842	-0.0838		-0.0774	-0.0756	-0.0745
		(0.0532)	(0.0535)	(0.0536)		(0.0523)	(0.0524)	(0.0524)
Economics Student		0.0864	0.0911*	0.0912*		0.0684	0.0730	0.0753
		(0.0522)	(0.0515)	(0.0516)		(0.0486)	(0.0484)	(0.0486)
Studied Game Theory		0.00541	0.00484	0.00535		0.0169	0.0158	0.0158
		(0.0481)	(0.0477)	(0.0478)		(0.0437)	(0.0437)	(0.0439)
Red Type			0.196***	0.174***			0.126***	0.127***
			(0.0235)	(0.0376)			(0.0229)	(0.0379)
Period			-0.0184***	-0.0136**			-0.0186***	-0.0122**
			(0.00339)	(0.00612)			(0.00324)	(0.00603)
Red x Mechanism				0.0413				-0.0579
				(0.0655)				(0.0640)
Red x Revelation				-0.0190				-0.00569
				(0.0577)				(0.0602)
Red x Punishment				0.0596				0.0491
				(0.0614)				(0.0574)
Period x Mechanism				-0.0280***				-0.0246***
				(0.00884)				(0.00863)
Period x Revelation				0.00204				-0.00448
				(0.00842)				(0.00904)
Period x Punishment				0.00440				0.00101
				(0.00827)				(0.00788)
Constant	0.207***	0.219***	0.231***	0.214***	0.227***	0.236***	0.280***	0.242***
	(0.0398)	(0.0574)	(0.0615)	(0.0713)	(0.0353)	(0.0524)	(0.0584)	(0.0670)
Observations	1920	1920	1920	1920	1848	1848	1848	1848
Adjusted R^2	0.029	0.048	0.108	0.113	0.035	0.053	0.094	0.097

Standard errors in parentheses
Standard errors clustered at group.
Last four columns drop observations outside [0,1]
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In the Revelation and Punishment conditions, participants are also asked to report their type, but only to their group members while making a voluntary contribution, allowing us to assess whether the ability to send messages about type can enable the group members themselves to enforce higher cooperation. The Revelation condition increases contributions over the baseline VCM by a very similar magnitude as Mechanism does. The difference is significant at the $p = .07$ level and at the $p < .03$ level when controls are included. However, in the Punishment condition, when participants can be held accountable for discrepancies between their report and their contribution, the positive effect of reporting disappears: the coefficient on Punishment is negative and insignificant across all specifications. Contributions in the Revelation condition, which is identical to Punishment except in the sanctioning phase, are significantly higher than Punishment at the $p < .05$ level across all specifications.

The additional models (2 - 8) reported in Table 1 confirm these main findings. The next three columns add controls for period, type, sex, and whether the participant has training in economics and game theory. As a further robustness check, columns (5) through (8) repeat the first four models dropping the 3.75% of observations in which the Percent Optimal Contribution variable is greater than 1 (which occurs if Red Types contributed more than 2) or less than zero (which occurs when Blue Types contributed less than 2). Dropping these observations slightly strengthens the difference between Mechanism and VCM and demonstrates that "mistakes" are not driving the main results.[8]

From Table 1, we also see that Period has a significant negative effect on contributions. In all four conditions, contributions decline significantly over time, although the effect is weaker in the Revelation condition.[9] Thus the ability to report one's type does not eliminate the standard decay in contributions over time. Finally, we note that Red Types contribute significantly more, as a percent of their optimal contribution, than Blue Types across all four conditions.

To better understand the source of the variation in contributions across conditions, we

[8] A two-limit Tobit model provides essentially identical results. In addition, if each group interacting across all ten periods is instead taken as the unit of observation, the coefficients on both Mechanism and Revelation are significant at at least the 6% level in every specification.

[9] Taking either the individual or the group as the unit of observation, the correlation between period and contribution is significant at the 1% level for Mechanism and the 5% level for VCM and Punishment. For Revelation, the correlation is significant at the 10% level if the individual is taken at the unit of observation and not significant at this level if the group is the unit of observation.

next consider the messages that participants sent in each of the three conditions for which messages were available: Mechanism, Revelation, and Punishment. Table 2 reports the relative frequency of messages sent by the Blue Types and the Red Types in each condition. In parentheses beside each report is the required contribution associated with it in the Mechanism condition. While reports of Purple, Yellow, and Green have different, specific contributions associated with them in the Mechanism condition, all are equally meaningless in the Revelation and Punishment conditions and are thus grouped together under the category "Nonsense."

Table 2: Frequency of Messages

Blue Types				Red Types			
Report	Mech.	Rev.	Punish.	Report	Mech.	Rev.	Punish.
Blue (4 in M)	.271	.694	.719	Red (2 in M)	.349	.774	.833
Purple (3 in M)	.094	-	-	Yellow (1 in M)	.139	-	-
Red (2 in M)	.545	.2	.235	Green (0 in M)	.472	-	-
Nonsense (R or P)	-	.106	.046	Nonsense (R or P)	-	.113	.056

Notes: This table reports the relative frequency of different messages by Blue Types and by Red Types in each condition. The number in parentheses gives the required contribution associated with the message in the Mechanism condition. "Nonsense" refers to any report of Purple, Yellow, or Green in the Revelation and Punishment conditions, where the messages do not have specific meaning.

The first row of the table shows the frequency of truthful reports in all three conditions: Blue Types who report Blue and Red Types who report Red. First, we see that, despite Mechanism's success at increasing contributions above the baseline VCM, untruthful reports dominate in this condition. Blue Types truthfully report in 27% of opportunities while Red Types truthfully report in 35% of opportunities.[10] In contrast, both types in the Revelation and Punishment conditions *typically report the truth*. The difference in truthtelling between Mechanism and Revelation or Mechanism and Punishment is significant at any reasonable level. Between one-fifth and one-quarter of Blue Types in these conditions do strategically report that they are Red Types and this is similarly common across the Revelation and Punishment conditions. The only difference in reports between these conditions is that participants in the Revelation condition are twice as likely to report one of the nonsense colors (Purple, Yellow, or Green) as those in the Punishment condition.[11] Therefore, the higher contributions achieved

[10]Note that the two types do not face a fully symmetric situation: While Blue Types can free ride by plausibly reporting that they are Red Types, Red Types can only free-ride by reporting that they are a type that others know does not exist. However, the difference in truthful messages across types is only significant at the $p = .075$ level if each individual-period is taken as the unit of observation, without controlling for repeated measures or session effects, and is otherwise not statistically significant.

[11]The difference in "Nonsense" reports is significant at the $p < .01$ level and at the $p = .074$ level if standard errors are clustered at the group-level.

under the Revelation condition do not appear to be the result of different messages being sent or, specifically, of Punishment participants being less honest.

Finally, we analyze the outcomes in the Punishment condition. Table 3 categorizes the possible Message-Contribution pairs and reports both the frequency and the average payoff reduction received in each case. First, participants could report a type and then make the optimal contribution for that type (Rows 1 and 2). In this case, the average reduction is small, close to two points for each type, but still significantly greater than zero.[12] Next, participants could report a type and then *under-contribute* relative to the social optimum for that type (Rows 3 and 4). In this case, they receive significantly higher payoff reductions, close to 4 points for each type. Since it costs an individual 5 points to increase his/her contribution by a token, this expected payoff reduction is *not sufficient* to change the incentives of someone who is under-contributing. In both of these two cases, the punishments are remarkably similar for those who report "Blue" vs. "Red," which suggests that group members are taking the reports as similarly credible.[13]

Finally, the last two rows show the reductions received after either reporting a type that does not exist or reporting a type that is mismatched with contribution – i.e., either reporting Blue and contributing less than 2 or reporting Red and contributing more than 2. In both cases, participants earn significantly higher payoff reductions, *even controlling for actual contributions*, than participants who do not submit a transparently false message.[14] Thus, while participants under-utilize the punishment mechanism as a tool for reducing free-riding, they appear *more* willing to punish blatant dishonesty. This provides further evidence that free-riding is not viewed as harshly as untruthful reporting.

[12] Clustering standard errors at the group level, reductions are significantly greater than zero at the 5% level for individuals reporting Red and contributing 2 and at the 10% level for individuals reporting Blue and contributing 4.

[13] In the case of participants reporting a type and contributing optimally for that report, those reporting Blue are telling the truth in 100% of observations while those reporting Red are telling the truth in less than 40% of observations.

[14] Participants who submit nonsense or mismatched reports have their earnings reduced by 2.9 more points ($p = .037$ or $p = .057$ with clustered standard errors) or by 2.72 points controlling for contribution ($p = .06$ or $p = .07$ with clustered standard errors).

Table 3: Punishment Summary Statistics by Contribution

Contribution/Report	Observations	Expected Reductions
Report Blue and Contribute Optimally	23	1.96
Report Red and Contribute Optimally	92	2.18
Report Blue and Under-contribute	223	4.09
Report Red and Under-contribute	172	4.01
Report Nonsense	27	6.88
Report-Contribution Mismatch	27	6.11

Notes: This table reports the frequency of possible message-contribution pairs in the Punishment condition and the expected payoff reductions associated each. *Report Nonsense* refers to a report of Purple, Yellow, or Green combined with any contribution. *Report-Contribution Mismatch* refers to participants who report Blue and contribute less than 2 or participants who report Red and contribute more than 2.

4 Discussion

This paper provides a novel test of preference revelation in a simple public goods game that was designed to emulate the standard public good experiments in the literature. To the best of my knowledge, this paper is the first to test whether "revealing demand" through a voluntary contribution is the same as "revealing demand" by reporting a type and whether simply requiring a report can promote cooperation. The results reported in this paper suggest that asking individuals to report their demand can lead to increased public good contributions. More specifically, the paper reports three main findings, discussed in detail below.

First, participants in the Mechanism condition contribute significantly more than participants in the VCM condition. In other words, asking participants to contribute by reporting a type, which transparently maps to a contribution level, leads to significantly higher contributions than simply asking people to voluntarily contribute. The question remains as to source of this difference. It is possible that the mechanism altered participants' beliefs about how much they were expected to contribute. Although such framing is a general feature of environments where participants are asked to report a demand, attempts were made to minimize this influence as much as possible, by emphasizing that behavior was confidential and that reporting a type different from one's own was completely acceptable and well within the rules of the game. While striking, this result is well-grounded in the literature on lying aversion across many different contexts. To further test the interpretation that the effect could be attributed to lying aversion, or a more general unwillingness to submit a report that is not true, we turn to the post-experiment questionnaire data. Participants in both conditions were asked the extent to

which they agreed that under-contributing (in the VCM) or under-reporting (in Mechanism) was "dishonest," for each of the two types. In both cases, individuals believed that making a sub-optimal contribution in the VCM condition was significantly *less dishonest* than making an advantageous misreport in the Mechanism condition ($Z = 3.275; p < .01$ for Blue Types and $Z = 3.006; p < .01$ for Red Types). Thus the questionnaire data do lend support to the interpretation that individuals do not view free-riding as being dishonest unless it requires them to underreport their demands and suggests that they do not view the contributions themselves as a revelation of demand.

Second, participants who are reporting to a binding mechanism send very different messages than those who are reporting only to their fellow group members. Participants who are reporting to their group typically report the truth ($70 - 83\%$ of observations). Further, under-reporting to group members is viewed as significantly more dishonest in questionnaire responses than under-reporting to a mechanism that will charge them (for Blue Types: $Z = 1.951; p < .1$ in Revelation and $Z = 2.359; p < .05$ in Punishment; for Red Types: $2.175; p < .05$ in Revelation and $Z = 2.335; p < .05$ in Punishment).

Finally, the honesty of group members' reports translates into higher contributions. Participants who are asked to report a type before voluntarily contributing, do contribute more. The magnitude of their contributions is very similar to those reporting to a binding mechanism. However, this effect goes away when group members can punish each other. Why? Participants are not any less honest in the Punishment condition: the messages sent were nearly identical. Instead, the poor performance of the Punishment mechanism appears to be the result of two related factors. First, the punishment mechanism is under-utilized, such that it is not in the interests of free-riders to increase their contributions to avoid punishment. This result is consistent with previous findings that punishment mechanisms are used less frequently when contributions are observed with noise (e.g. Ambrus and Greiner 2012) and further suggests that participants are wary of punishing individuals when it isn't clear who amongst them is most deserving. Rather than harshly punishing individuals who under-contribute relative to their reported type, participants instead tend to punish obviously untruthful reports. Second, the presence of the punishment mechanism appears to undermine the positive effect of truthfully

reporting one's type that was present in the Revelation condition. The backfiring of punishment has been well-documented in gift-exchange games when punishment is not sufficient to incentivize participants to cooperate (Fehr, Gächter, and Kirchsteiger 1997; Aquino, Gazzale, and Jacobson 2015) and the result is also consistent with the findings of Gneezy and Rustichini (2000) that weak fines can crowd out cooperative behavior. The questionnaire data again lend further credence to this interpretation. Specifically, participants in the Punishment condition, who can be held accountable for their contributions, do not view under-contributing after making their report as being as dishonest as participants in the Revelation condition do (Blue Types: $Z = 2.36; p < .05$; Red Types: $Z = 2.6; p < .01$).

The findings of this experiment have implications both for how economists think about problems of demand revelation and for how to effectively solicit contributions in the field from individuals with unobservable preferences. The experiment finds that asking participants to report their demand-type, either to their group members or to a mechanism, can generate higher contributions. Further, the results indicate that participants view misreporting their demand when directly asked as being less acceptable than free-riding when making a voluntary contribution: Misreporting one's type occurs less frequently than under-contributing in an otherwise identical game, is considered more dishonest in questionnaire responses, and is punished more harshly in the incentivized experiment. Theoretically, these results suggest that the design of mechanisms could be enhanced by taking advantage of agents' natural inclination to tell the truth. In addition, framing truthful revelation as such – a natural requirement in cases where agents are assumed to be aware of their own preference parameters – could improve the success of truthtelling mechanisms in the lab. There are also clear applications to promoting cooperation in groups or raising contributions in the field. The results suggest that communities or organizations raising money for a project or cause could generate significantly more revenue by asking individuals to self-identify as being of a specific beneficiary type. Additional work is needed to determine the extent to which such an approach could improve contributions or provision in practice.

5 References

Ambrus, Attila and Ben Greiner (2012). "Imperfect Public Monitoring with Costly Punishment: An Experimental Study." *The American Economic Review*, 102(7): 3317-3332.

Aquino, Patrick, Robert Gazzale, and Sarah Jacobson (2015). "When Do Punishment Institutions Work?"

Attiyeh, Greg, Robert Franciosi, and R. Mark Isaac (2000). "Experiments with the Pivot Process for Providing Public Goods." *Public Choice*, 102:95-114.

Baldry, J.C. (1986) "Tax evasion is not a gamble." *Economics Letters*, 22: 333 - 335.

Bornstein, Gary, and Ori Weisel (2010). "Punishment, Cooperation, and Cheater Detection in 'Noisy' Social Exchange." *Games*, 1(1): 18-33.

Cappelen, Alexander W., Erik Ø. Sørensen, and Bertil Tungodden (2013). "When Do We Lie?" *Journal of Economic Behavior & Organization*, Vol. 93: 258-265.

Cason, Timothy N., Tatsuyoshi Saijo, Tomas Sjöström, and Takehiko Yamato (2006). "Secure implementation experiments: Do strategy-proof mechanisms really work?" *Games and Economic Behavior*, 57(2): 206-235.

Chan, Kenneth S., Stuart Mestelman, Robert Moir, and R. Andrew Muller (1999). "Heterogeneity and the Voluntary Provision of Public Goods." *Experimental Economics*, Vol. 2: 5-30.

Charness, Gary and Martin Dufwenberg (2006). "Promises and Partnership." *Econometrica*, 74(6): 1579-1601.

Charness, Gary and Martin Dufwenberg (2011). "Participation." *American Economic Review*, 101(4): 1211-1237.

Chen, Yan (2008). "Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Literature." In *The Handbook of Experimental Economics Results*, edited by Charles R. Plott and Vernon Smith. North-Holland.

Chen, Yan and Charles R. Plott (1996). "The Groves-Ledyard Mechanism: An Experimental Study of Institutional Design." *Journal of Public Economics*, 59: 335-364.

Cohn, Alain, Ernst Fehr, and Michel André Maréchal (2014). "Business culture and dishonesty in the banking industry." *Nature*, 516, 86-89.

Deneckere, Raymond and Sergei Severinov (2008). "Mechanism design with partial state verifiability." *Games and Economic Behavior*, 64: 487-513.

Ellingsen, Tore, Magnus Johannesson, Jannie Lilja and Henrik Zetterqvist (2009). "Trust and

Truth." *The Economic Journal*, 119: 252-276.

Erat, Sanjiv and Uri Gneezy (2012). "White Lies." *Management Science*, Vol. 58 (4): 723-733.

Falkinger, Josef, Ernst Fehr, Simon Gächter, and Rudolf Winter-Ebmer (2000). "A Simple Mechanism for the Efficient Provision of Public Goods: Experimental Evidence." *The American Economic Review*, Vol. 90(1): 247-264.

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

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

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

Fischbacher, Urs and and Franziska Föllmi-Heusi (2013). "Lies in Disguise." *Journal of the European Economic Association*, 11(3): 525547.

Gneezy, Uri (2005). "Deception: The Role of Consequences." *The American Economic Review*, Vol. 95 (1), 384-394.

Gneezy, Uri and Aldo Rustichini (2000). "A Fine is a Price." *The Journal of Legal Studies*, 29: 117.

Green, Jerry and Jean-Jacques Laffont (1977). "Characterization of Strongly Individually Incentive Compatible Mechanisms for the Revelation of Preferences for Public Goods." *Econometrica* Vol. 45: 427438.

Green, Jerry and Jean-Jacques Laffont (1986). "Partially Verifiable Information and Mechanism Design." *Review of Economic Studies*, 53: 447456.

Healy, Paul J. (2006). "Learning Dynamics for Mechanism Design: An Experimental Comparison of Public Goods Mechanisms." *Journal of Economic Theory*, 129: 114 -149.

Isaac, R. Mark and James M. Walker (1988). "Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism. *Economic Inquiry*, Vol. 26(4): 585-608.

Keser, Claudia (1996). "Voluntary Contributions to a Public Good When Partial Contribution is a Dominant Strategy." *Economics Letters*, Vol. 50: 359-366.

Krajbich, Ian, Colin Camerer, John Ledyard, and Antonio Rangel (2009). "Using Neural Measures of Economic Value to Solve the Public Goods Free-Rider Problem." *Science*, Vol. 326: 596-599.

Kawagoe, Toshiji and Toru Mori (2001). "Can the Pivotal Mechanism Induce Truth-telling? An Experimental Study." *Public Choice*,108: 331354

Ledyard, John O. (1995). "Public Goods: A Survey of Experimental Research." In *The Handbook of Experimental Economics*, edited by A.E. Roth and J. Kagel. Princeton University Press.

Lundquist, Tobias, Tore Ellingsen, and Magnus Johannesson (2009). "The Aversion to Lying." *Journal of Economic Behavior & Organization*, Vol. 70: 81-92.

Mazar, Nina, On Amir, and Dan Ariely (2008). "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research*, 45: 633-644.

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

Robbett, Andrea (2014). "Sustaining Cooperation in Heterogeneous Groups" *Working Paper.*

Rondeau, Daniel, William D. Schulze, and Gregory L. Poe (1999). "Voluntary Revelation of the Demand for Public Goods Using a Provision Point Mechanism." *Journal of Public Economics*, 72: 455-470.

Samuelson, Paul (1954). "A Pure Theory of Public Expenditure." *The Review of Economics and Statistics*, Vol. 36(4): 387-389.

Appendix: Supplemental Figures

Table 4: Distribution of Contributions

	Red Types					Blue Types			
Contribution	VCM	Mech.	Rev.	Punish.	Contribution	VCM	Mech.	Rev.	Punish.
0	131	92	95	136	0	4	13	2	11
1	27	27	27	51	1	9	10	6	8
2	49	88	73	47	2	203	139	161	237
3	0	4	0	0	3	25	24	18	26
4	1	4	0	0	4	31	69	68	24
Mean	0.620	0.980	0.887	0.620	Mean	2.257	2.494	2.565	2.144

Notes: This table reports the frequency of contributions by Red and Blue Types in each condition.

Appendix: Instructions

Instructions

Hello and welcome to our experiment. Please follow along with these instructions as I read them aloud.

Payment and Confidentiality

For your participation today, you have already earned $5. You will earn an additional amount of money that depends on the number of *points* you accumulate in the experiment. It is therefore important that you understand the instructions. Please raise your hand if you have questions at anytime. Please also know that we never deceive participants in economics experiments – these instructions are an accurate description of how the experiment will proceed and we will not tell you anything untrue or misleading.

The experiment consists of two parts. You can think of Part 1 as a practice round to enable you to get accustomed to the experiment. After both parts have been completed, you will be paid, in cash, your point total from Part 2. The exchange rate between points and dollars is:

50 points = $1.00.

In this experiment, your decisions will be confidential; none of the other participants will ever know the decisions you make.

Instructions for Part 1

In this part, you will be randomly matched with two other participants in the room to form a group of three people. You will interact with this same group for several periods.

In each period, each individual will receive an endowment of 4 tokens. Each group member will decide how many of these tokens to contribute to a group account. The more tokens contributed to the group account, the more points each person in the group earns for the period. However, the individual who made the contribution *could* earn fewer points as a result.

The exact number of points that you will earn from the contributions made by you and your group members will depend on the payoff table assigned to you. There are two types of payoff tables, which we'll call Blue Type and Red Type. They will each be described in detail in a moment.

At the end of each period, you will learn: the total contributions in your group and your payoff for the period. The contribution of each individual group member will also be displayed in separate boxes on the screen, as shown below:

Figure 3: Part 1 Instructions Page 1 of 2 (Distributed in All Conditions)

The order of these boxes is randomly generated each period, and so no one will be able to track any individual's contributions across periods.

For the first 5 periods of Part 1, everyone will have **Red Type payoffs**, which are shown in the table below:

Points for Red Type

Total Contributions of Other Two Group Members \ Own Contribution	0	1	2	3	4
0	10	5	0	0	0
1	20	15	10	0	0
2	30	25	20	0	0
3	40	35	30	0	0
4	50	45	40	0	0
5	60	55	50	0	0
6	70	65	60	0	0
7	80	75	70	0	0
8	90	85	80	0	0

Interpretation of Table: The columns are labeled 0 through 4 along the top of the table and correspond to the tokens that *you* could contribute. The rows are labeled 0 through 8 along the left side of the table and correspond to the contributions that the *other two* members could contribute in total. The cells show the points that you would receive in each case. For instance: Imagine that everyone in your group contributed 2. To find your payoff for the period, we would look in the column marked "2" for your own contribution and in the row marked "4" for the total contributions of the two others. We find that your payoff for the period is 40 points.

Please note:

- As you increase your own contribution from 0 to 1 token or from 1 to 2 tokens, your points decrease by 5 (we can see this by looking from left to right in the table). If you contribute 3 or 4 tokens, you will receive zero points for the period.
- For each extra token contributed by someone, the other two group members receive an additional 10 points each (we can see this by looking from top to bottom in the table).

For the next 5 periods of Part 1, you will be re-matched into a new group and everyone will have **Blue Type payoffs**, which are shown in the table below:

Points for Blue Type

Total Contributions of Other Two Group Members \ Own Contribution	0	1	2	3	4
0	20	25	30	25	20
1	30	35	40	35	30
2	40	45	50	45	40
3	50	55	60	55	50
4	60	65	70	65	60
5	70	75	80	75	70
6	80	85	90	85	80
7	90	95	100	95	90
8	100	105	110	105	100

Please note:

- As you increase your own contribution from 0 to 1 token or from 1 to 2 tokens, your points *increase* by 5. As you increase your own contribution from 2 to 3 tokens or 3 to 4 tokens, your points *decrease* by 5.
- As before, for each extra token contributed by someone, the other two group members receive an additional 10 points each.

Figure 4: Part 1 Instructions Page 2 of 2 (Distributed in All Conditions)

R

Instructions for Part 2

We will now begin Part 2. You will be paid according to the total points you accumulate in this part. Part 2 differs slightly from Part 1, so please follow along with the instructions carefully. Part 2 will last for 10 periods. You will be re-matched with two different participants to form a new group of 3 people and you will be in this same group for all 10 periods.

Not everyone in your group will have the same payment structure. At the start of *each period*, each individual will be randomly assigned to be either a Blue Type or a Red Type, each with equal probability. You will learn your own type at the start of the period but not the types of the two other individuals in your group. Any composition is possible – thus, the other two group members could both be Blue Types, they could both be Red Types, or one could be Blue and the other Red. The types will be re-assigned in every period and no one will ever learn which type each person was assigned.

Remember that the points for the two types are as follows:

Points for Blue Type

Total Contributions of Other Two Group Members \ Own Contribution	0	1	2	3	4
0	20	25	30	25	20
1	30	35	40	35	30
2	40	45	50	45	40
3	50	55	60	55	50
4	60	65	70	65	60
5	70	75	80	75	70
6	80	85	90	85	80
7	90	95	100	95	90
8	100	105	110	105	100

Points for Red Type

Total Contributions of Other Two Group Members \ Own Contribution	0	1	2	3	4
0	10	5	0	0	0
1	20	15	10	0	0
2	30	25	20	0	0
3	40	35	30	0	0
4	50	45	40	0	0
5	60	55	50	0	0
6	70	65	60	0	0
7	80	75	70	0	0
8	90	85	80	0	0

Sometimes participants find it useful to send messages about their payoff types to each other. At the start of each period, you will learn your assigned type for the period. You will then be asked your type and will enter a message to your group members: you can enter "I am a Blue Type" or "I am a Red Type." You can also enter "I am a Green Type," "I am a Yellow Type," or "I am a Purple Type." You are permitted to be untruthful in these messages. Your group members will see your message and your contribution, but never your actual type.

Each person will see the messages sent by the group members *only after* making their contributions for the period.

Just as in Part 1, each period you will learn the total contributions in your group and your payoff for the period. The contribution of each individual group member will also be displayed in separate boxes on the screen – *along with the message this person sent* about his/her type. The order of these boxes is randomly generated each period, and so no one will be able to track any individual's contributions across periods.

Figure 5: Revelation Condition Page 1 of 2 (Note that VCM is the same, but omits mentions of messages.)

R

To summarize, in each period you will:

- Learn your type
- Submit a message about your type to your group members
- Make a contribution
- Learn the contributions and messages of each of your group members

Figure 6: Revelation Condition Page 2 of 2 (Note that VCM is the same, but omits mentions of messages.)

P

Instructions for Part 2

We will now begin Part 2. You will be paid according to the total points you accumulate in this part. Part 2 differs slightly from Part 1, so please follow along with the instructions carefully. Part 2 will last for 10 periods. You will be re-matched with two different participants to form a new group of 3 people and you will be in this same group for all 10 periods.

Not everyone in your group will have the same payment structure. At the start of *each period*, each individual will be randomly assigned to be either a Blue Type or a Red Type, each with equal probability. You will learn your own type at the start of the period but not the types of the two other individuals in your group. Any composition is possible – thus, the other two group members could both be Blue Types, they could both be Red Types, or one could be Blue and the other Red. The types will be re-assigned in every period and no one will ever learn which type each person was assigned.

Remember that the points for the two types are as follows:

Points for Blue Type

Total Contributions of Other Two Group Members	Own Contribution 0	1	2	3	4
0	20	25	30	25	20
1	30	35	40	35	30
2	40	45	50	45	40
3	50	55	60	55	50
4	60	65	70	65	60
5	70	75	80	75	70
6	80	85	90	85	80
7	90	95	100	95	90
8	100	105	110	105	100

Points for Red Type

Total Contributions of Other Two Group Members	Own Contribution 0	1	2	3	4
0	10	5	0	0	0
1	20	15	10	0	0
2	30	25	20	0	0
3	40	35	30	0	0
4	50	45	40	0	0
5	60	55	50	0	0
6	70	65	60	0	0
7	80	75	70	0	0
8	90	85	80	0	0

Sometimes participants find it useful to send messages about their payoff types to each other. At the start of each period, you will learn your assigned type for the period. You will then be asked your type and will enter a message to your group members: you can enter "I am a Blue Type" or "I am a Red Type." You can also enter "I am a Green Type," "I am a Yellow Type," or "I am a Purple Type." You are permitted to be untruthful in these messages. Your group members will see your message and your contribution, but never your actual type.

Each person will see the messages sent by the group members *only after* making their contributions for the period. After everyone has made their contribution, you will learn the tokens contributed by each of the group members as well as each of their messages. You will then have an opportunity to reduce the point earnings of one or both of your group members for the period. It costs you a point to reduce another group member's point total by 3 points.

Just as in Part 1, each period you will learn the total contributions in your group and your payoff for the period. The contribution of each individual group member will also be displayed in separate boxes on the screen – *along with the message this person sent* about his/her type. The order of these boxes is randomly generated each period, and so no one will be able to track any individual's contributions across periods.

Below each box, you will be able to enter how many points, if any, you'd like to pay to reduce this person's point total. This person will then have their points reduced by *three times* that

Figure 7: Punishment Condition Page 1 of 2

amount (plus any reductions from the other group member). For instance, imagine you choose to pay 1 point to reduce someone's earnings and the other group member pays 2 points to reduce that person's earnings. The individual's earnings would then be reduced by (1+2) *times* 3 = 9 points. Finally, you will see the amount by which each individual's earnings was reduced for the period (in this example: 9).

Regardless of what the group members pay, an individual will never have his/her earnings for a given period reduced below 0. Note that it is possible to earn a negative payoff if you spend more reducing someone's earnings than you earn in the period. However, you can always avoid this outcome by making different decisions. If your point total over all ten periods is less than zero, this amount will be subtracted from your participation payment.

To summarize, in each period you will:

- Learn your type
- Submit a message about your type to your group members
- Make a contribution
- Learn the contributions and messages of each of your group members and decide whether to reduce their earnings
- Learn the contributions, messages, and reductions of each of your group members

Figure 8: Punishment Condition Page 2 of 2

M

Instructions for Part 2

We will now begin Part 2. You will be paid according to the total points you accumulate in this part. Part 2 differs slightly from Part 1, so please follow along with the instructions carefully. Part 2 will last for 10 periods. You will be re-matched with two different participants to form a new group of 3 people and you will be in this same group for all 10 periods.

Not everyone in your group will have the same payment structure. At the start of *each period*, each individual will be randomly assigned to be either a Blue Type or a Red Type, each with equal probability. You will learn your own type at the start of the period but not the types of the two other individuals in your group. Any composition is possible – thus, the other two group members could both be Blue Types, they could both be Red Types, or one could be Blue and the other Red. The types will be re-assigned in every period and no one will ever learn which type each person was assigned.

Remember that the points for the two types are as follows:

Points for Blue Type

Total Contributions of Other Two Group Members	Own Contribution 0	1	2	3	4
0	20	25	30	25	20
1	30	35	40	35	30
2	40	45	50	45	40
3	50	55	60	55	50
4	60	65	70	65	60
5	70	75	80	75	70
6	80	85	90	85	80
7	90	95	100	95	90
8	100	105	110	105	100

Points for Red Type

Total Contributions of Other Two Group Members	Own Contribution 0	1	2	3	4
0	10	5	0	0	0
1	20	15	10	0	0
2	30	25	20	0	0
3	40	35	30	0	0
4	50	45	40	0	0
5	60	55	50	0	0
6	70	65	60	0	0
7	80	75	70	0	0
8	90	85	80	0	0

In this part, you will not directly choose how many tokens to contribute to the group account. Instead, the computer will ask you what your type is and then will automatically deduct a certain number of tokens as your contribution, based on what you tell it. If you report that you are a Blue Type, the computer will automatically contribute 4 tokens for you. If you report that you are a Red Type, the computer will automatically contribute 2 tokens for you. You are permitted to be untruthful in these messages. The table below shows the possible messages that you can send the computer when it asks your type, as well as the number of tokens that you will then automatically contribute.

Message	Tokens You Will Contribute
I am a Red Type	2
I am a Blue Type	4
I am a Green Type	0
I am a Yellow Type	1
I am a Purple Type	3

Your group members will see your message, but never your actual type.

Just as in Part 1, each period you will learn the total contributions in your group and your payoff for the period. The contribution of each individual group member will also be displayed in separate boxes on the screen along with the message this person sent about his/her type. The order of these boxes is randomly generated each period, and so no one will be able to track any individual's contributions across periods.

Figure 9: Mechanism Condition Page 1 of 2

M

To summarize, in each period you will:

- Learn your type
- Submit a message about your type that determines your contribution
- Learn the contributions and messages of each of your group members

Figure 10: Mechanism Condition Page 2 of 2