Public Good Provision Mechanisms and Reciprocity

DOI:
10.1016/j.jebo.2019.02.001

Document Version
Accepted author manuscript

Public Good Provision Mechanisms and Reciprocity

Maria Kozlovskaya[a] and **Antonio Nicolò**[b]

[a]Aston University, Birmingham, UK. m.kozlovskaya@aston.ac.uk (corresponding author)

[b]University of Padua, Italy and University of Manchester, UK. antonio.nicolo@unipd.it

January 2, 2019

Abstract

This paper determines optimal public good provision mechanisms in an environment where agents are motivated by reciprocity. In a two-agent economy, we show that the standard pivot mechanism is not strategy-proof if at least one agent cares strongly enough about reciprocity. Truthful reporting maximises a player's own payoff and hence is viewed as selfish by his opponent, who retaliates by understating demand for the public good. Incentive compatibility is restored if the mechanism is implemented sequentially. When agents are asked to report their demands in turn, a high report by the first mover (he) may result in him paying an unfairly large share of the provision cost, should the second mover (she) understate her demand. Hence, truthful reporting is not judged as selfish by the second mover, who reciprocates by stating her true demand. Our results alert the institutional designer to the importance of game dynamics when agents are non-selfish.

Keywords: Psychological Game Theory, Mechanism Design, Reciprocity

JEL classification: C79, D63, D82, H41.

1 Introduction

Optimal provision of public goods relies on incentive-compatible mechanisms. It is well known that markets and voluntary donation schemes under-supply public goods owing

to a free-rider problem. One solution to this problem is taxation which makes donation compulsory. Although taxation eliminates free-riding, it does not solve the designer's information problem: an optimal level of provision depends on agents' private valuations of the public good, which they may be willing to understate.

More complex provision schemes, known as mechanisms, can incentivise agents to report their valuations truthfully. Among the class of mechanisms that are strategy-proof (truthful reporting is a dominant strategy for each agent), individually rational (every agent weakly prefers participating in the mechanism than opting out) and efficient (for each outcome, there is no other outcome weakly preferred by all participants), the pivot mechanism plays a central role: in a standard environment, it has the smallest budget deficit.

In this paper we investigate whether the pivot mechanism maintains its desirable properties when agents are endowed with other-regarding preferences, which play a central role in public-good provision settings (Andreoni, 1990; Ledyard, 1995; Fischbacher et al., 2001; Page et al., 2005). Public projects are often managed by communities, whose members care about each other's actions and welfare. As a result, these projects attract significant voluntary donations, contrary to the theoretical prediction of complete free-riding: In 2017, individuals donated £10.3 billion in the UK (Charitable Aid Foundation UK, 2018) and $286.65 billion in the US (Giving USA Foundation, 2018). Although some of this generosity is attributable to image concerns, even in anonymous laboratory settings people contribute between 40 and 60% of the social optimum (Ledyard, 1995; Chaudhuri, 2011). Experiments indicate that the most relevant type of other-regarding preferences in public good provision is reciprocity – a desire to sacrifice one's own payoff in order to match others' contributions (Keser and Van Winden, 2000; Croson, 2007; Chaudhuri, 2011). We model reciprocity following Dufwenberg and Kirchsteiger (2004), and adopt a psychological game theory framework (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009); as a solution concept we employ the perfect Bayesian equilibrium of a psychological game developed by Attanasi et al. (2016) (see also Section 6.2 in Battigalli and Dufwenberg (2009)) suitably adapted for reciprocal preferences.

We show that the desirable properties of a mechanism depend on revelation dynamics when agents are reciprocal. This is in contrast to selfish preferences, where the pivot mechanism is strategy-proof regardless of the way it is implemented. Reciprocity constitutes an enlargement of the preference domain; hence one would expect that the set

of incentive-compatible mechanisms reduces. Our paper confirms this intuition: the pivot mechanism is not strategy-proof if agents care enough about reciprocity. Reciprocal agents are willing to sacrifice their payoff to punish unkind actions. An action is considered unkind if it yields a smaller-than-average payoff for the opponent, where the average is taken across all Pareto-efficient strategies. Importantly, strategies involving understatement of demand are Pareto-inefficient, since they harm both the opponent and the player himself, as guaranteed by the incentive compatibility of the pivot mechanism for the selfish preferences case. Truthful reporting therefore earns a below-average payoff for the opponent among Pareto-efficient strategies, and is assessed as unkind. A reciprocal player retaliates by understating her demand, even though it reduces her material payoff. Bierbrauer and Netzer (2016) first observed this feature of incentive-compatible mechanisms under reciprocity. Their proposed solution is to offer players subsidies which are constructed to remove the dependence between their payoffs and hence to cut off the retaliation channel. However, as Bierbrauer and Netzer (2016) themselves emphasise, subsidising players is not always possible, especially if the mechanism designer is unable to cover the ex-post deficit.

We present a costless way to restore incentive-compatibility by asking the players to report their valuations in turn. In a dynamic mechanism, reports can be conditioned on earlier play, which expands the strategy sets and alters the perception of truthful reporting. In particular, late movers can understate their demand after observing high reports. This means that an early mover reporting high valuation may be sacrificing their payoff for the benefit of others. It follows that truthful reporting is not viewed as unkind, and will indeed be a mutual best response regardless of reciprocity sensitivity.

Although our framework can be readily extended to countable player sets, in this paper we focus on two-player mechanisms. First, it allows us to compare our findings to existing results on implementability under reciprocal preferences (Bierbrauer and Netzer, 2016). Second, as the number of players increases, the dimensionality of the higher-order belief hierarchy grows exponentially. Indeed, a player is required to hold beliefs on what each of his opponents thinks about everyone else's strategy, which is cognitively challenging. Hence a realistic theory of reciprocity in n-player games should draw on models of rational inattention to incorporate information processing costs, which lies beyond the scope of this paper. Instead, we focus on providing a sharp robustness check for mechanism design theory, as well as highlighting the importance of revelation dynamics when agents

are non-selfish.

The rest of this paper proceeds as follows. Related literature is discussed in Section 2. Section 3 models the pivot mechanism under reciprocity as a psychological game of incomplete information. Subsection 3.1 describes payoffs and preferences, while subsection 3.2 introduces the equilibrium concept we use. Subsections 3.3 and 3.4 solve the simultaneous and sequential reporting cases, respectively. We discuss Pareto-inferior equilibria in Subsection 3.5 and provide concluding remarks in Section 4. The Appendix contains the proofs.

2 Literature

Formal models of reciprocity were developed by Rabin (1993) (for static games) and Dufwenberg and Kirchsteiger (2004) (for dynamic games). In these papers reciprocity is defined as a desire to reward good intentions and punish bad intentions. Experimental research has demonstrated the importance of intention-based preferences in public good provision settings. Croson (2007) reports laboratory results which strongly support reciprocity over other types of other-regarding preferences in public goods games. In line with this finding, a contemporary meta-study (Chaudhuri, 2011) emphasises the role of beliefs in the game: players contribute more if they think their opponents will be generous. One of the most recent experiments on public goods to date (Dhami et al., 2018) confirmed that reciprocity is a significant motivation in this game. This large and growing body of empirical evidence justifies our choice of a specific model of preference representation, to which we commit in our analysis for reasons of clarity.

In this paper we use one of the most comprehensive formal models of reciprocity (Dufwenberg and Kirchsteiger, 2004). Reciprocal preferences imply that utility depends on higher-order beliefs, because an opponent's behaviour can be interpreted differently depending on his expectations of others' behaviour. Did the person who issued me a last-minute wedding invitation know that I am going on holiday on that day? My belief about his belief (*i.e.* my second-order belief) crucially affects my judgement of his kindness. In a strategic setting, games with belief-dependent utility are modelled in the framework of Psychological Game Theory (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009). Bayesian analysis of dynamic psychological games was pioneered in Attanasi et al. (2016), whose solution concept we employ. Importantly, psychological games are a strict

extension of standard games, hence reciprocal preferences include selfish preferences as a special case. Naturally, a mechanism designer's task is more challenging in the larger preference domain.

To the best of our knowledge, the only other model of mechanism design with intention-based preferences is Bierbrauer and Netzer (2016), who characterise incentive-compatible static mechanisms under preferences for reciprocity. Our paper offers a complementary perspective by focusing on public goods provision and extending the analysis to dynamic mechanisms. Several authors have incorporated other types of other-regarding preferences into mechanism design. Kucuksenel (2012) characterizes the class of interim efficient mechanisms for agents with altruistic preferences, whereas Tang and Sandholm (2012) study optimal auctions with spiteful bidders. While altruism and spitefulness are generally empirically relevant concerns, in public goods provision settings reciprocity has been shown to play a more important role (Croson, 2007).

Finally, our paper also contributes to the growing literature looking at strategy-proof mechanisms in extensive form games. Bergemann and Välimäki (2010) generalise the pivot mechanism for the case when agents learn their valuations over time. Li (2017) introduces a refinement of strategy-proofness, called "obviously strategy-proofness", that extends this concept to extensive form games. In another recent paper Schummer and Velez (2017) show in a standard preference domain that under certain conditions, when preference revelation is sequential an agent could even have a strict disincentive to truthfully report preferences even though outcomes are computed using a strategy-proof social choice function. In line with this strand of research, our paper highlights the importance of game form and determines optimal revelation dynamics for the case of reciprocal preferences.

3 The Pivot Mechanism under Reciprocal Preferences

3.1 The Environment

Consider two agents who contemplate producing an indivisible non-excludable public good. The agents differ in their valuations of the public good, which are denoted θ_i for

$i = 1, 2$ and are independently uniformly distributed over $\Theta_i \equiv \{\theta_L, \theta_H\}$.[1] The cost of the good is c. To rule out trivial cases, we assume $\theta_L + \theta_H > c > 2\theta_L$. The vector of types is denoted $\theta \equiv (\theta_1, \theta_2) \in \Theta \equiv \Theta_1 \times \Theta_2$.

A mechanism is a pair (M, g), where $M = M_1 \times M_2$ is a cross product of message spaces and $g : M \to X$ is an outcome function (or game form) mapping into the set of outcomes $X = \{0, 1\} \times R^2$. An outcome specifies whether the public good is produced (1) or not (0) and assigns a fee (transfer) that each each agent pays (receives).

In a direct mechanism the message space coincides with the set of reported types: $M = \Theta_1 \times \Theta_2$. In our environment a direct mechanism can be written as a decision rule $q : \Theta \to \{0, 1\}$ (where $q(\theta) = 1$ means a good is produced) and a transfer rule $t_i : \Theta \to R$ for $i = 1, 2$ (denoting a fee paid by the agent to the community).

An agent's strategy in a direct mechanism is a function which maps his true type into his reported type: $s_i : \Theta_i \to \Theta_i$. The strategy set is denoted S_i. We restrict the attention to truthful mechanisms so we can focus on pure strategies without loss of generality. Agent i's initial belief about j's strategy (first-order belief) is a probability distribution over S_j, denoted $\alpha_i \in \Delta(S_j)$ where $i \neq j$. In particular, $\alpha(s_j) \in [0, 1]$ is the probability which i assigns to j playing strategy s_j, where $\sum_{s_j \in S_j} \alpha(s_j) = 1$. When modelling reciprocity we will also use second-order beliefs. Agent i's initial belief about j's belief about i's strategy is denoted $\beta_i \in \Delta(S_i)$.[2]

In a mechanism with a sequential game form, Player 1 (he) is assumed to publicly report his valuation first. Player 2 (she) can condition her report not only on her own type, but also on Player 1's report, hence her strategy is $s_2 : \Theta_2 \times \Theta_1 \to \Theta_2$. Players 2 may also

[1] Our choice of type domain implies a high degree of symmetry in valuations – both between agents and within the type space for each agent. The symmetry assumption is not wlog; however, it complements the definition of the reference point in the reciprocity model we use. Dufwenberg and Kirchsteiger (2004) define an action as kind if it results in a higher-than-equitable payoff for the opponent, where the equitable payoff is a midpoint between the maximum and minimum payoffs a player can get. In order for the equitable payoff to adequately reflect a morally neutral outcome and to be useful for judging about relative kindness of the two players, the space of possible payoffs should be symmetrical between agents and around that reference point, for which our assumptions on the type space are necessary. A countable type space is used for clarity of exposition. Our results extend to continuous type space unless c is too large.

[2] Note that both first- and second-order beliefs in our paper are modelled as probability distributions over a player's strategy set. More generally, k-order beliefs in psychological game theory are probability distributions over a product of the opponents' strategy set and $k-1$-order-belief set (Geanakoplos et al., 1989). In dynamic games, beliefs are conditional probability systems assigning a probability distribution over strategy and belief sets to each possible history (Battigalli and Dufwenberg, 2009). They way we define β corresponds to a *feature* of this more general second-order belief system, namely, the expectation of the opponent's belief about the player's own strategy.

update her belief about Player 1's strategy after observing his report. Correspondingly, Player 1 may update his belief about Player 2's belief about his strategy.[3] We denote i's updated beliefs as $\alpha_i^K \in \Delta(S_j)$ and $\beta_i^K \in \Delta(S_i)$ respectively, where $K \in \{L, H\}$ corresponds to the first mover's report (low or high).

Finally, an agent's payoff from taking part in the mechanism $\pi_i(\theta; \theta_i)$ depends on the vector of reported types θ (which determines the provision decision and transfers) and his true type θ_i (potentially different from his reported type; determines his enjoyment from the public good if produced). We follow the literature and assume a quasilinear environment with linear valuations, where an agent's payoff is the difference between his enjoyment of the public good $\theta_i q(\theta)$ and the monetary transfer t_i:

$$\pi_i(\theta; \theta_i) = \theta_i q(\theta) - t_i(\theta). \tag{1}$$

A mechanism is a game of incomplete information; hence best responses will be calculated in terms of expected payoffs.[4] An agent's ex-ante expected payoff from taking part in a mechanism $\langle q(\cdot), t(\cdot) \rangle$ and following a strategy s_i, given his belief about j's strategy α_i is given by:

$$\Pi_i(s_i, \alpha_i) = \mathrm{E}_{\theta_i, \theta_j, s_j} \left[\theta_i q\left(s_i\left(\theta_i\right), s_j\left(\theta_j\right)\right) - t_i\left(s_i\left(\theta_i\right), s_j\left(\theta_j\right)\right) \right]. \tag{2}$$

The opponent's expected payoff ($\Pi_j(s_i, \alpha_i)$) is calculated analogously to (2). For a reciprocal player, it is also important what the opponent expects their payoffs to be, as we explain in more detail in the next subsection. The opponent's expectation of the player's payoff is taken with respect to both the first- and the second-order beliefs (that is, treating the player's own strategy as a random variable over which he holds a probabilistic second-order belief β_i):

$$\Pi_i(\beta_i, \alpha_i) = \mathrm{E}_{\theta_i, \theta_j, s_j, s_i} \left[\theta_i q\left(s_i\left(\theta_i\right), s_j\left(\theta_j\right)\right) - t_i\left(s_i\left(\theta_i\right), s_j\left(\theta_j\right)\right) \right]. \tag{3}$$

[3] Beliefs about Player 2's strategy are updated in a trivial way: $\alpha_1^K = \alpha_1$ and $\beta_2^K = \beta_2$, because her strategy is already conditioned on Player 1's report.

[4] $\Pi_i(s_i, \alpha_i)$ is the ex-ante payoff, *i.e.* calculated before the types are realised. The expectation is taken with respect to both players' types, which are random variables at that point, and Player i's probabilistic first-order belief α_i. Using ex-ante expected payoffs is standard in the mechanism design literature. It is straightforward to show that a strategy which maximises ex-ante expected payoff of Player i maps each type Player i into a report which maximises that type's interim expected payoff (the payoff calculated after the player knows his own type but treating the other player's type as a random variable).

3.1.1 Kindness Function

A reciprocal agent's utility U_i is not identical to his "material" payoff Π_i. In addition to Π_i it includes a "reciprocity component" $Y_i\kappa_i\lambda_i$ which captures his enjoyment from rewarding kind actions, and punishing unkind ones:

$$U_i\left(s_i, \alpha_i, \beta_i\right) = \Pi_i\left(s_i, \alpha_i\right) + Y_i\kappa_i\left(s_i, \alpha_i\right)\lambda_i\left(\beta_i, \alpha_i\right), \tag{4}$$

where $Y_i \in R_+$ is i's sensitivity to reciprocity, $\kappa_i\left(s_i, \alpha_i\right)$ is i's kindness to j and $\lambda_i\left(\beta_i, \alpha_i\right)$ is i's belief about j's kindness to i, as explained below.

The utility model (4) was introduced in the seminal paper on sequential reciprocity by Dufwenberg and Kirchsteiger (2004). The paper defines i's kindness $\kappa_i\left(s_i, \alpha_i\right)$ as the extra payoff which i's strategy yields for j on top of j's "equitable" payoff $\Pi_j^e(\alpha_i)$:

$$\kappa_i(s_i, \alpha_i) = \Pi_j(s_i, \alpha_i) - \Pi_j^e(\alpha_i). \tag{5}$$

$\Pi_j^e(\alpha_i)$ is a reference point: the payoff j "deserves" and is computed as the average between the highest and the lowest payoffs i can yield for j, by varying his strategy:

$$\Pi_j^e\left(\alpha_i\right) = \frac{1}{2} \cdot \left[\max\left\{\Pi_j\left(s_i, \alpha_i\right) \mid s_i \in S_i^E\right\} + \min\left\{\Pi_j\left(s_i, \alpha_i\right) \mid s_i \in S_i^E\right\}\right]. \tag{6}$$

Importantly, only Pareto-efficient strategies of the player ($S_i^E \subset S_i$) are taken into account when calculating his opponent's equitable payoff. A strategy is *not* efficient if it always results in worse payoffs for all players than some other strategy, no matter what others do (Dufwenberg and Kirchsteiger, 2004).[5]

Turning to Player i's belief about his opponent's kindness λ_i, it is defined analogously to i's own kindness:

$$\lambda_i(\beta_i, \alpha_i) = \Pi_i(\beta_i, \alpha_i) - \Pi_i^e(\beta_i). \tag{7}$$

An important implication of the utility model (4) is that the belief that j is kind ($\lambda_i > 0$) makes Player i's utility increasing in his own kindness, and vice versa.

[5]Formally, a strategy $s_i \in S_i$ is efficient if there does not exist another strategy $s_i' \in S_i$ such that for all strategies of the opponent $s_j \in S_j$ it holds that $\Pi_k(s_i', s_j) \geq \Pi_k(s_i, s_j)$ for both players $k \in \{1, 2\}$, with strict inequality for at least one of them.

3.2 Equilibrium Concept

The pivot mechanism with simultaneous revelation is a static game of incomplete information. The appropriate equilibrium concept should rely on Bayesian analysis, which was first used in incomplete-information psychological games by Attanasi et al. (2016) and Bierbrauer and Netzer (2016). Our definition follows their approach.

A Bayesian equilibrium of a simultaneous psychological game is a strategy profile $s^* = (s_1^*, s_2^*)$ such that

(a) (s_1^*, s_2^*) maximises both players' expected utility given their initial beliefs $(\alpha_1, \beta_1, \alpha_2, \beta_2)$;

(b) beliefs are consistent, *i.e.* derived from strategies: $\beta_i(s_i^*) = \alpha_j(s_i^*) = 1$ for $i, j \in \{1, 2\}$ and $i \neq j$.

In a dynamic mechanism, an equilibrium should satisfy the principle of perfection: later movers should maximise their utility conditional on updated beliefs. Moreover, a dynamic consistency requirement for beliefs needs to specify how updated beliefs are formed. The weakest rationality principle would require that updated beliefs α_i^K and β_i^K are *compatible with the observed history of play*, *i.e.* are not ruled out by Player 1's report θ_K.[6] In other words, α_2^K (or β_1^K) is compatible with θ_K if it assigns positive probabilities only to strategies which map at least one type θ_i, $i \in \{L, H\}$ into the observed report θ_K. For example, Player 2's belief that Player 1 may be always reporting a low demand $(\alpha_2^K(s^L) > 0)$ is not compatible with a high report of Player 1.[7]

Wherever we can, we would like to impose stricter rationality requirements on beliefs. Along the equilibrium path, updated beliefs can be derived from initial beliefs using the Bayes rule. This is how updated belief consistency is modelled in Attanasi et al. (2016), who were the first to study intention-based preferences in an dynamic Bayesian game. Our equilibrium definition borrows from their paper.

A perfect Bayesian equilibrium of a dynamic psychological game is a strategy profile $s^* = (s_1^*, s_2^*)$ such that

(a) (s_1^*, s_2^*) maximises Player 1's expected utility given his initial beliefs (α_1, β_1) and Player 2's expected utility given her updated beliefs $(\alpha_2^K, \beta_2^K, K \in \{L, H\})$;

[6] In the terminology of Battigalli and Dufwenberg (2009), strategies α_2^K and β_1^K should *allow* history θ_K.

[7] Note that this is a weak requirement. In the space of all possible beliefs $(\Delta(S_i) \subset [0, 1]^4)$, the set of beliefs not compatible with any given report has measure zero.

(b) beliefs are consistent:

(i) initial beliefs are derived from strategies: $\beta_i(s_i^*) = \alpha_j(s_i^*) = 1$ for $i, j \in \{1, 2\}$ and $i \neq j$;

(ii) updated beliefs are derived from initial beliefs using Bayes rule whenever possible, and are compatible with the observed history of play.[8]

3.3 Implementation with a Simultaneous Game Form

The pivot mechanism consists of an efficient decision rule, where the public good is produced if and only if the sum of agents' valuations exceed its cost, and a transfer rule which aligns agents' incentives with social welfare. Formally, the decision rule is given by:

$$q = 1 \text{ iff } \theta_1 + \theta_2 \geq c, \tag{8}$$

while the transfer rule is defined by:

$$t_i(\theta) = \theta_L q\left(\theta_L, \theta_j\right) + \left(q\left(\theta\right) - q\left(\theta_L, \theta_j\right)\right)\left(c - \theta_j\right). \tag{9}$$

According to (9), an agent's payment does not depend on his report, unless it is pivotal, in which case the payment imposed on agent i equals the change in other agents' welfare caused by his report.

Proposition 1 *Truthful reporting is not a Bayesian equilibrium of the pivot mechanism with simultaneous reporting if at least one agent cares enough about reciprocity.*

Sketch of Proof: Observe that, according to the transfer rule (9), a player's payment is weakly decreasing – and hence payoff weakly increasing – in the opponent's report. It follows that the maximum (minimum) possible expected payoff the player can get, holding his own strategy fixed, is obtained when his opponent always reports a high (low)

[8]We require beliefs to be compatible with the observed history because Bayes rule does not rule out any updated beliefs following off-equilibrium play, including the ones which contradict that play. Our equilibrium notion can be viewed as a pure-strategy-equivalent of sequential equlibrium in Kreps and Wilson (1982) and Battigalli and Dufwenberg (2009, p. 27). Note that we cannot directly use their definitions because they rely on converging sequences of fully mixed strategies, whereas our model focuses on pure strategies. If we considered mixed strategies, Battigalli and Dufwenberg (2009) consistency definition would rule out updated beliefs which contradict the observed history of play, hence the compatibility requirement would not be necessary.

demand. Note, however, that always reporting a low demand is an inefficient strategy: compared to truthful reporting, it reduces both players' payoffs. Hence truthful reporting by the opponent earns the player the lowest possible payoff among efficient strategies, which means it is assessed as unkind. A player's utility is decreasing in the unkind opponent's payoff, according to the model (4). Hence, for a high enough reciprocity sensitivity Y_i, the player would like to deviate from truthful reporting (which maximises the material component of the utility) to always reporting a low demand (which minimises the opponent's payoff and hence maximises the reciprocity component of the utility). In other words, if at least one player is sufficiently reciprocity-sensitive, truthful reporting will not be a mutual best response. The formal proof is available in the Appendix.

Proposition 1 shows that the mechanism with the smallest budget deficit of all individually rational and incentive-compatible (under selfish preferences) schemes is not strategy-proof when agents care enough about each other's intentions. In the terminology of Bierbrauer and Netzer (2016), the pivot does not have the "insurance property": by changing their reports, agents can affect each others' payoffs. Hence, a player retaliates by understating his demand if he thinks his opponent's reporting strategy is "unkind".

It is rather obvious that a strategy-proof mechanism which implements efficient outcomes in a certain environment may not be so in a larger preference domain. The standard method of fine-tuning a mechanism to a new environment is altering the transfer rule, usually by providing additional subsidies to the players (Bose et al., 2006; Bodoh-Creed, 2012; Bierbrauer and Netzer, 2016). However, this may be infeasible for a budget-constrained social planner, especially in the case of the pivot mechanism which is popular precisely because it achieves the smallest budget deficit of all incentive compatible individually rational mechanisms.

Instead of manipulating the transfers, we propose to change the dynamics of revelation. Hence, our result could offer a costless solution to the "negative reciprocity trap". Importantly for the mechanism designer, reciprocal feelings are menu-dependent. In other words, the same strategy can be judged as kind or unkind depending on what other strategies were available to the player. This feature of the model is crucial for our result. Moving to a dynamic game allows players to condition their report on the history of past moves and hence expands their strategy sets. In the next section we explain how this expansion alters the kindness perception of truthful reporting.

3.4 Implementation with a Sequential Game Form

Proposition 2 *Truthful reporting is a perfect Bayesian equilibrium of the pivot mechanism with sequential reporting regardless of the agents' reciprocity sensitivity.*

Proof: in the Appendix.

In a dynamic environment the report of the first mover (he) is observed by the second mover (she). A high report makes the first mover vulnerable to free-riding: the second mover can understate her demand and make the first mover pay an unfairly high share of the provision cost.[9] This means understating by the first mover is not an inefficient strategy any more, because it protects him from a free-riding second mover. As a result, truthful reporting is now judged as neutral in terms of kindness, being kinder than understatement but less kind than overstatement. This means that a reciprocity-concerned mover doesn't feel the need to retaliate, and truthful reporting is implemented in equilibrium regardless of the player's reciprocity sensitivity.

3.5 Spiteful Equilibrium

Games with intention-based preferences often have more equilibria then games with selfish preferences. In particular, reciprocity gives rise to a "negative coordination" equilibrium, where both players choose a Pareto-dominated strategy out of spite. It turns out that this problem persists in the simultaneous pivot mechanism. If agents care enough about reciprocity, the pivot mechanism has an equilibrium where agents report low demand for the public good, irrespective of their true type. In this equilibrium both agents retaliate to an "unkind" action of their opponent. Hence for the simultaneous pivot we can make a stronger prediction than just lack of truthful reporting (Proposition 1). A strategy profile where both agents always report a low demand (hereafter *understating*) will be an equilibrium of the game:

Proposition 3 *Understating is a Bayesian equilibrium of the pivot mechanism with simultaneous reporting if both players care enough about reciprocity.*

[9] Under the transfer rule of the pivot mechanism, the second mover has nothing to gain from understating her demand in this case, but the availability of this conditional strategy alters the set of efficient strategies and hence the equitable payoff.

Proof: in the Appendix.

Similarly to the previous section, implementation with a sequential game form overturns this negative result. As Proposition 4 shows, always reporting a low demand is not an equilibrium of a dynamic game.

Proposition 4 *Understating is not a perfect Bayesian equilibrium of the pivot mechanism with sequential reporting.*

Proof: in the Appendix.

The intuition behind the proof is as follows. In a strategy profile where both players always report low demand, an unexpected high report by the first mover would make his opponent update her first-order belief. Although the Bayes rule does not pin down the exact updated belief in this case, in order to be compatible with the observed history it has to assign probability zero to understating. Player 2 will now assign a positive probability to at least one of the following strategies of Player 1: reporting truthfully (which is neutral in terms of kindness), always reporting the opposite of his true demand (which is also neutral) or to always reporting a high demand (which is kind). On average, Player 1 is being weakly kind to Player 2 in this off-equilibrium history. Hence, Player 2 would like to deviate to truthful reporting, which is less unkind than understating and also weakly increases her material payoff.

4 Discussion and Conclusion

The feelings of reciprocity increase donations in voluntary public good provision schemes. Our paper shows that the same feelings may have the opposite effect in strategy-proof mechanisms, such as the pivot. Participants in the mechanism view each other's truthful reports and subsequent payments as selfish, because this behaviour is materially rewarded. Reciprocity-sensitive agents retaliate "selfishness" by understating their own demand, which leads to under-provision of the public good.

In a dynamic setting a high report by the first mover may make him vulnerable to free-riding by the second mover. Always reporting a low demand would protect the first mover from free-riding at the other player's expense. Choosing to report truthfully means sacrificing such protection, which is not selfish and hence not viewed as unkind. This fact enables us to restore incentive compatibility by implementing the mechanism sequentially.

Our finding highlights the importance of game structure in fostering public good contributions – a fact which has been experimentally established by Andreoni and Samuelson (2006) for Prisoner's Dilemma (a.k.a. a voluntary public good provision mechanism).

Although we considered a special case of a simple mechanism in a 2-player economy, some of our results hold more generally. In particular, the difficulty in truthful implementation under reciprocal preferences highlighted by Proposition 1 extends to all strategy-proof mechanisms. By construction, truthful reporting maximises a player's expected payoff in these mechanisms. Any deviation from it would either hurt both players (hence being inefficient), or increase the opponent's payoff while decreasing the player's own payoff. Hence, of all Pareto-efficient strategies, truthful reporting is the least kind. This result provides a mechanism-design explanation for the well-documented crowding out effect between material and moral incentives (Frey and Oberholzer-Gee, 1997; James Jr, 2005; Holmås et al., 2010).

Our findings concerning sequential revelation also illustrate a more general insight. Compared to static games, dynamic games have extended strategy sets. A strategy which was at the boundary of the payoff efficiency frontier in a static setting (hence viewed as unkind as possible) will likely be at the interior of the frontier in a dynamic setting, and consequently viewed as more kind. Truthful reporting is such a boundary strategy in simultaneous mechanisms, owing to incentive compatibility. Sequential revelation weakens the dominance of truthful reporting and consequently its assessment as extremely selfish. From an institutional designer's point of view, revealing early choices makes the pioneers look brave and inspires followers to respond in kind.

Appendix

Proof of Proposition 1.

Before proceeding to the proof we introduce useful notation. Let s_i^T denote a truthful reporting strategy of Player i, so that $s_i^T(\theta_i) = \theta_i$ for all θ_i. In addition, denote a strategy where the player always reports low demand by $s_i^L : s(\theta_k) = \theta_L$ for $k \in \{H, L\}$, and one where he always reports high demand by $s_i^H : s(\theta_k) = \theta_H$ for $k =\in \{H, L\}$. Finally, if the player reports the opposite of his true demand we denote it by $s^{-T} : s_i^{-T}(\theta_k) = \theta_j$ for $k, j \in \{H, L\}$ and $k \neq j$.

Consider Player i's expected material payoff if both himself and the other player report truthfully:

$$\Pi_i(s_i^T, s_j^T) = \frac{0}{4} + \frac{\theta_H + \theta_L - c}{4} + \frac{\theta_L - \theta_L}{4} + \frac{\theta_H - \theta_L}{4} = \frac{2\theta_H - c}{4}. \tag{10}$$

The player's payoff in all other strategy profiles can be calculated analogously. The message game can be written down in normal form (Table 1).

Table 1: Payoff Table of Simultaneous Pivot Mechanism

		s_2			
		s^T	s^{-T}	s^L	s^H
s_1	s^T	$\frac{2\theta_H-c}{4}$, $\frac{2\theta_H-c}{4}$	$\frac{2\theta_H-c}{4}$, $\frac{\theta_H+\theta_L-c}{4}$	$\frac{\theta_H+\theta_L-c}{2}$, $\frac{\theta_H-\theta_L}{4}$	$\frac{\theta_H-\theta_L}{2}$, $\frac{\theta_H+\theta_L-c}{2}$
	s^{-T}	$\frac{\theta_H+\theta_L-c}{4}$, $\frac{2\theta_H-c}{4}$	$\frac{\theta_H+\theta_L-c}{4}$, $\frac{\theta_H+\theta_L-c}{4}$	$\frac{2\theta_L-c}{2}$, $\frac{\theta_H-\theta_L}{4}$	$\frac{\theta_H-\theta_L}{2}$, $\frac{\theta_H+\theta_L-c}{2}$
	s^L	$\frac{\theta_H-\theta_L}{4}$, $\frac{\theta_H+\theta_L-c}{2}$	$\frac{\theta_H-\theta_L}{4}$, $\frac{2\theta_L-c}{2}$	0, 0	$\frac{\theta_H-\theta_L}{2}$, $\frac{\theta_H+2\theta_L-2c}{2}$
	s^H	$\frac{\theta_H+\theta_L-c}{2}$, $\frac{\theta_H-\theta_L}{2}$	$\frac{\theta_H+\theta_L-c}{2}$, $\frac{\theta_H-\theta_L}{2}$	$\frac{\theta_H+2\theta_L-2c}{2}$, $\frac{\theta_H-\theta_L}{2}$	$\frac{\theta_H-\theta_L}{2}$, $\frac{\theta_H-\theta_L}{2}$

It can be shown by examining the payoff matrix that s^{-T} and s^L are inefficient strategies, since s^T guarantees a weakly larger payoff for both players than either s^{-T} or s^L, for any strategy of the opponent. In contrast, strategies s^T and s^H are efficient.

Consider a strategy profile where both players report their demands truthfully. Is it a mutual best response? In order to answer this question, players' kindness needs to be calculated. By a slight abuse of notation, we will write $\alpha_i = s_j$ and $\beta_i = s_i$ to denote i's deterministic beliefs, that is, his belief that j plays s_j and believes s_i with probability 1.

If Player i reports truthfully, and believes that j reports truthfully ($\alpha_i = s^T$), and believes that j believes that he (i) reports truthfully ($\beta_i = s^T$), then i's judgement about j's kindness is

$$\lambda_i(s^T, s^T) = \Pi_i(s^T, s^T) - \Pi_i^e(s^T) = \frac{2\theta_H - c}{4} - \frac{1}{2} \cdot \left(\frac{2\theta_H - c}{4} + \frac{\theta_H - \theta_L}{2} \right) = \frac{2\theta_L - c}{8} < 0. \tag{11}$$

where $\Pi_i^e(s^T)$ is i's equitable payoff, calculated as the average between his highest and lowest possible payoffs, assuming j is playing an efficient strategy (s^T or s^H) and i himself is reporting truthfully. A truth-telling Player i's highest possible payoff is realised when Player j always reports a high demand ($\max\left\{\Pi_i\left(s^T, \alpha_i\right) \mid \alpha_i \in \{s^T, s^H\}\right\} = \Pi_i\left(s^T, s^H\right) = \frac{\theta_H - \theta_L}{2}$), while his lowest possible payoff on the efficiency frontier is realised when Player j reports truthfully ($\min\left\{\Pi_i\left(s^T, \alpha_i\right) \mid \alpha_i \in \{s^T, s^H\}\right\} = \Pi_i\left(s^T, s^T\right) = \frac{2\theta_H - c}{4}$).

It follows that a truth-telling Player j is being unkind to i. Player i can either respond by truth-

telling, which maximises his material payoff, or retaliate by always reporting a low demand, which maximises the reciprocity component of his utility. The player's reciprocity sensitivity determines which of the two strategies is the best response.

Responding with $s_i = s^L$ while believing that j reports truthfully ($\alpha_i = s^T$) is as unkind as possible:

$$\kappa_i(s^L, s^T) = \Pi_j(s^L, s^T) - \Pi_j^e(s^T) = \frac{\theta_H + \theta_L - c}{2} - \frac{4\theta_H - 2\theta_L - c}{8} = \frac{3\,(2\theta_L - c)}{8} < \frac{2\theta_L - c}{8} < 0. \quad (12)$$

Player i's overall utility from playing s^T (while believing that j plays s^T and believes s^T) equals

$$U_i\left(s^T, s^T, s^T\right) = \Pi_i\left(s^T, s^T\right) + Y_i \kappa_i\left(s^T, s^T\right) \lambda_i\left(s^T, s^T\right) = \frac{2\theta_H - c}{4} + Y_i \cdot \frac{(2\theta_L - c)^2}{64}, \quad (13)$$

while his utility from playing s^L (while believing that j plays s^T and believes s^T) equals

$$U_i\left(s^L, s^T, s^T\right) = \frac{\theta_H - \theta_L}{4} + Y_i \cdot \frac{3\,(2\theta_L - c)^2}{64}. \quad (14)$$

It is straightforward to show that (14) is larger than (13) iff

$$Y_i > \frac{8\,(\theta_H + \theta_L - c)}{(2\theta_L - c)^2}. \quad (15)$$

Hence, if at least one of the players is sensitive enough to reciprocity, truthful reporting is not an equilibrium, *i.e.* the pivot mechanism is not strategy-proof, *q.e.d.*

Proof of Proposition 2.

If the mechanism is implemented sequentially, Player 2 can condition her move on Player 1's report. Player 2 now has 16 conditional strategies at her disposal: $s_2(\theta_1) : \{\theta_L, \theta_H\} \times \{\theta_L, \theta_H\} \to \{\theta_L, \theta_H\}$, while Player 1 still has four. Denote Player 2's strategy by two superscripts, according to the notation introduced in the proof of Proposition 1. *E.g.* s^{-TT} denotes the strategy of Player 2 who lies following a low report by Player 1 and tells the truth following a high report.

Will truthful reporting be an equilibrium in this game? In order to find out a player's best response to truthful reporting, we need to determine that player's assessments of the opponent's kindness. First, consider Player 2's assessment of Player 1's kindness. The range of possible payoffs Player 2 can get if

she always reports truthfully is the same as in the simultaneous case (since this strategy of Player 2 is not conditioned on Player 1's report, her expected payoffs are presented in the first column of Table 1). Her highest possible payoff is $\Pi_2\left(s^H, s^{TT}\right) = \frac{\theta_H - \theta_L}{2}$ and her lowest possible payoff is $\Pi_2\left(s^L, s^{TT}\right) = \frac{\theta_H + \theta_L - c}{2}$.

Note that in the dynamic game s_1^L is not an inefficient strategy any more. In particular, it earns Player 1 a higher payoff than truthful reporting when Player 2 plays s^{HL}, that is, responds to a low report by always reporting high and responds to a high report by always reporting low:

$$\Pi_1\left(s^L, s^{HL}\right) = \frac{\theta_H - \theta_L}{2} > \frac{\theta_H + \theta_L - c}{2} = \Pi_1\left(s^T, s^{HL}\right). \tag{16}$$

It follows that s_1^L should be taken into account for the sake of Player 2's equitable payoff calculation:

$$\Pi_2^e(s^{TT}) = \frac{1}{2} \cdot \left(\frac{\theta_H - \theta_L}{2} + \frac{\theta_H + \theta_L - c}{2}\right) = \frac{2\theta_H - c}{4}, \tag{17}$$

which is exactly equal to Player 2's actual payoff if both players report truthfully. This means Player 1's kindness to Player 2 is zero:

$$\lambda_2(s^{TT}, s^T) = \Pi_2(s^{TT}, s^T) - \Pi_2^e(s^{TT}) = \frac{2\theta_H - c}{4} - \frac{2\theta_H - c}{4} = 0. \tag{18}$$

Whenever at least one of the kindness terms is zero, the player's overall utility is equal to her material payoff, which is maximised by always reporting truthfully. Hence Player 2's best response to truthful reporting is also truthful reporting.

Turning to Player 1's choice, note that the highest possible payoff he can get if he reports truthfully is $\frac{\theta_H - \theta_L}{2}$ (*e.g.* if Player 2 plays s^{HH}) and the lowest possible payoff is $\frac{\theta_H + \theta_L - c}{2}$ (*e.g.* if Player 2 plays s_2^{HL}).

Also observe that s_2^{HL} is not an inefficient strategy of Player 2, because it earns her opponent a higher payoff than truthful reporting, when Player 1 himself plays s_1^{-T}:

$$\Pi_1\left(s^{-T}, s_2^{HL}\right) = \frac{\theta_H + \theta_L - c}{2} > \frac{\theta_H + \theta_L - c}{4} = \Pi_1\left(s^{-T}, s_2^{TT}\right). \tag{19}$$

It follows that s_2^{HL} should be taken into account for the sake of Player 1's equitable payoff calculation:

$$\Pi_1^e(s^T) = \frac{1}{2} \cdot \left(\frac{\theta_H - \theta_L}{2} + \frac{\theta_H + \theta_L - c}{2} \right) = \frac{2\theta_H - c}{4}, \tag{20}$$

which is exactly equal to Player 1's actual payoff if both players report truthfully. This means Player 2's kindness to Player 1 is zero:

$$\lambda_1(s^T, s^{TT}) = \Pi_1(s^T, s^{TT}) - \Pi_1^e(s^T) = \frac{2\theta_H - c}{4} - \frac{2\theta_H - c}{4} = 0. \tag{21}$$

This means Player 1's overall utility equals his material payoff, which is maximised by always reporting truthfully. Hence truthful reporting is a mutual best response, *q.e.d.*

Proof of Proposition 3.

Consider a strategy profile where both players always report low demand: $(s_1, s_2) = (s^L, s^L)$. We will now show that this profile can be an equilibrium of the game.

Recall that in a simultaneous case s^L and s^{-T} are inefficient strategies. Hence a player's equitable payoff if he always reports low equals

$$\Pi_i^e(s^L) = \frac{1}{2} \left(\Pi_i(s_i^L, s_j^T) + \Pi_i(s_i^L, s_j^H) \right) = \frac{1}{2} \cdot \left(\frac{\theta_H - \theta_L}{4} + \frac{\theta_H - \theta_L}{2} \right) = \frac{3\theta_H - 3\theta_L}{8}. \tag{22}$$

Player i's judgement about j's kindness is then

$$\lambda_i(\beta_i, \alpha_i) = \Pi_i(s^L, s^L) - \Pi_i^e(s^L) = 0 - \frac{3\theta_H - 3\theta_L}{8} = \frac{3\theta_L - 3\theta_H}{8} < 0. \tag{23}$$

If Player i deviates by reporting truthfully his kindness to j will be

$$\kappa_i(s^T, s^L) = \Pi_j(s^T, s^L) - \Pi_j^e(s^L) = \frac{\theta_H - \theta_L}{4} - \frac{3\theta_H - 3\theta_L}{8} = \frac{\theta_L - \theta_H}{8} > \frac{3\theta_L - 3\theta_H}{8}. \tag{24}$$

The condition below ensures that such deviation is unprofitable:

$$U_i \left(s^L, s^L, s^L \right) \geq U_i \left(s^T, s^L, s^L \right), \tag{25}$$

which can be expanded as

$$0 + Y_i \cdot \frac{9\left(\theta_L - \theta_H\right)^2}{64} \geq \frac{\theta_H + \theta_L - c}{2} + Y_i \cdot \frac{3\left(\theta_L - \theta_H\right)^2}{64}, \tag{26}$$

which holds iff

$$Y_i \geq \frac{16\left(\theta_H + \theta_L - c\right)}{3\left(\theta_H - \theta_L\right)^2}. \tag{27}$$

Hence, if $Y_i \geq \frac{16(\theta_H+\theta_L-c)}{3(\theta_H-\theta_L)^2}$ for both players, then understating is an equilibrium, *q.e.d.*

Proof of Proposition 4.

We will now show that understating (*i.e.* both players reporting low demand irrespective of type and any observed history) is not an equilibrium of a sequential pivot. We do it by demonstrating that, in a strategy profile (s_1^L, s_2^{LL}), the second mover's strategy s_2^{LL} is not a best response to an updated belief after a high report by the first mover, hence (s_1^L, s_2^{LL}) is not an equilibrium.

Consider a strategy profile (s_1^L, s_2^{LL}). In an off-equilibrium history where Player 1 reports θ_H, Player 2 updates her first-order belief to a strategy which is compatible with such a high report, *i.e.* $\alpha_2^H(s^L) = 0$, whereas $\alpha_2^H(s_2) > 0$ for $s_2 \in \{s^T, s^{-T}, s^H\}$.

Let us determine Player 1's kindness according to such an updated belief of Player 2.

Player 2's highest possible payoff if she plays s^{LL} is $\Pi_2\left(s^H, s_2^{LL}\right) = \frac{\theta_H - \theta_L}{2}$ and her lowest possible payoff is $\Pi_2\left(s^L, s_2^{LL}\right) = 0$.

Recall that in the dynamic game s^L is not an inefficient strategy any more, as established in the proof of Proposition 2. Hence it should be taken into account for the purpose of equitable payoff calculation:

$$\Pi_2^e(s^{LL}) = \frac{1}{2} \cdot \left(\frac{\theta_H - \theta_L}{2} + 0\right) = \frac{\theta_H - \theta_L}{4}, \tag{28}$$

which is exactly equal to Player 2's actual payoff if Player 1 reports truthfully, or always lies. If Player 1 always reports a high demand, Player 2's payoff is above the equitable level. In summary, all strategies of Player 1 in the support of Player 2's updated first-order belief give rise to a payoff of Player 2 which is greater or equal than $\Pi_2^e(s^{LL})$. This means their convex combination is also at least as large as the equitable payoff *i.e.* Player 1's kindness to Player 2 is non-negative:

$$\lambda_2(s^{LL}, \alpha_2^H) = \alpha_2^H(s^T)\left(\frac{\theta_H - \theta_L}{4}\right) + \alpha_2^H(s^{-T})\left(\frac{\theta_H - \theta_L}{4}\right) + \alpha_2^H(s^H)\left(\frac{\theta_H - \theta_L}{2}\right) - \frac{\theta_H - \theta_L}{4} \geq 0. \tag{29}$$

If Player 1's kindness is zero, then Player 2's best response is to report truthfully, which maximises her material payoff equal to her overall utility in this case. Hence a deviation from s^{LL} to s^{TT} is profitable.

If Player 1's kindness is positive, Player 2 can weakly increase the material component of her utility and strictly increase the reciprocity component of her utility by deviating to truthful reporting. Indeed, s^{TT} earns her opponent a strictly larger payoff than s^{LL} regardless of his action and hence is less unkind.

Hence s^{LL} is not a best response to $\alpha_2^H = s^H$. We have shown that $\left(s^L, s^{LL}\right)$ is not an equilibrium, because s^{LL} is not a best response to any updated first-order belief α_2^H compatible with observed history, *q.e.d.*

References

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464–477.

Andreoni, J. and L. Samuelson (2006). Building rational cooperation. *Journal of Economic Theory 127*(1), 117–154.

Attanasi, G., P. Battigalli, and E. Manzoni (2016). Incomplete-information models of guilt aversion in the trust game. *Management Science 62*(3), 648–667.

Battigalli, P. and M. Dufwenberg (2009). Dynamic psychological games. *Journal of Economic Theory 144*(1), 1–35.

Bergemann, D. and J. Välimäki (2010). The dynamic pivot mechanism. *Econometrica 78*(2), 771–789.

Bierbrauer, F. and N. Netzer (2016). Mechanism design and intentions. *Journal of Economic Theory 163*, 557–603.

Bodoh-Creed, A. L. (2012). Ambiguous beliefs and mechanism design. *Games and Economic Behavior 75*(2), 518–537.

Bose, S., E. Ozdenoren, and A. Pape (2006). Optimal auctions with ambiguity. *Theoretical Economics 1*(4), 411–438.

Charitable Aid Foundation UK (2018). CAF UK Giving 2018: An overview of charitable giving in the UK.

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

Croson, R. T. (2007). Theories of commitment, altruism and reciprocity: evidence from linear public goods games. *Economic Inquiry 45*(2), 199–216.

Dhami, S., M. Wei, and A. al Nowaihi (2018). Public goods games and psychological utility: Theory and evidence. *Journal of Economic Behavior and Organization*.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*(2), 268–98.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters 71*(3), 397–404.

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review 87*(4), 746–755.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological Games and Sequential Rationality. *Games and Economic Behavior 1*(1), 60–79.

Giving USA Foundation (2018). Giving USA 2018: The annual report on philanthropy for the year 2017.

Holmås, T. H., E. Kjerstad, H. Lurås, and O. R. Straume (2010). Does monetary punishment crowd out pro-social motivation? A natural experiment on hospital length of stay. *Journal of Economic Behavior & Organization 75*(2), 261–267.

James Jr, H. S. (2005). Why did you do that? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance. *Journal of Economic Psychology 26*(4), 549–566.

Keser, C. and F. Van Winden (2000). Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics 102*(1), 23–39.

Kreps, D. and R. Wilson (1982). Sequential Equilibrium. *Econometrica* (50), 863–894.

Kucuksenel, S. (2012). Behavioral mechanism design. *Journal of Public Economic Theory 14*(5), 767–789.

Ledyard, J. (1995). Public goods: some experimental results. In J. Kagel and A. Roth (Eds.), *Handbook of Experimental Economics*. Princeton: Princeton University Press.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review 107*(11), 3257–87.

Page, T., L. Putterman, and B. Unel (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *The Economic Journal 115*(506), 1032–1053.

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

Schummer, J. and R. A. Velez (2017). Sequential preference revelation in incomplete information settings. *Working paper*.

Tang, P. and T. Sandholm (2012). Optimal auctions for spiteful bidders. In *AAAI*.