Vickrey-Clarke-Groves Mechanism and Preference for Reciprocity

Maria Kozlovskaya* and Antonio Nicolò †

February 28, 2017

Abstract

This paper applies psychological game theory to mechanism design. We study an environment where agents care about reciprocity, and show that the Vickrey-Clarke-Groves mechanism is not incentive compatible under such preferences. However, incentive compatibility is restored if the mechanism is implemented sequentially. We consider a 2-player sequential pivot mechanism with reciprocity-concerned players and prove that true reporting is the only equilibrium in this case.

Keywords: Psychological Game Theory, Mechanism Design, Reciprocity

JEL classification: C79, D63, D82, H41.

1 Introduction

In a Public Good Provision game, each agent is better off free-riding on his opponents (i.e. contributing zero amount to the public good) irrespective of what others do; hence, zero contributions by everyone is the only Nash Equilibrium of the game. This prediction is in sharp contrast with laboratory observations. A survey on experimental results in public goods conducted by Ledyard (1995) documents the expected contribution between 40 and 60% of the total optimum.

*University of Huddersfield. Email: m.kozlovskaya@hud.ac.uk (corresponding author and the source of all mistakes in the paper).
†University of Padua and University of Manchester.
Moving from voluntary contributions to more complex institutions involving communication (i.e. mechanisms) guarantees public good provision even in a theoretical equilibrium. Importantly, this equilibrium relies on the assumption that agents only care about their material well-being. However, robust institution design needs to rely on more realistic models of human behaviour, accounting for “non-selfish”, or “social” preferences.

There have been numerous attempts to incorporate non-selfish preferences into mechanism design. Kucuksenel (2012) characterizes the class of interim efficient mechanisms for agents with altruistic preferences (an altruist’s utility function depends not only on their own material well-being but also on the average of other agents’ well-beings). Although such an interpretation is able to explain some registered phenomena of more-than-selfish contribution, the assumption that agents are altruistic towards others regardless of their behaviour is incompatible with experimental evidence. In particular, Ledyard (1995) concludes, by carefully analyzing the design and results of many public good experiments, that when the conflict between group interest and self-interest is removed, subjects still contribute in ways that are counter to both their self-interest and their group interest.

Models of reciprocal behaviour, where people tend to be altruistic towards only those who are good with them, provide a better approximation of experimental outcomes in the public good provision game. The most parsimonious model of reciprocity to date is developed in Dufwenberg and Kirchsteiger (2004). Their paper incorporates desirable features of earlier reciprocity models (Rabin, 1993) and extends them to a strictly wider framework of dynamic games. For this reason, we adopt their approach as the main utility model in this paper, and evaluate the performance of VCG mechanism under preference for reciprocity as defined in Dufwenberg and Kirchsteiger (2004).

Our first result is negative: we show that the standard VCG mechanism is not incentive compatible when agents care about reciprocity. In particular, there will be an equilibrium where agents report low demand for the public good, irrespective of their true type. This is because, in such a report profile, both agents consider their opponent’s under-reporting “unkind” (in Dufwenberg and Kirchsteiger (2004) terminology), and best-respond by also under-reporting, hence retaliating the opponent’s unkind action.

We then show that incentive compatibility is restored if the mechanism is implemented sequentially, i.e. by asking the agents to report their demand for the public good in turn. In the unique equilibrium of this reporting game, both agents report their true type. The intuition here is as follows. In the last stage of the game, the second player will retaliate
the first mover’s action, which results in both players having a high payoff (if the action was true reporting) or low payoff (is the action was under-reporting). Knowing this, the first mover will report truthfully.

In order to solve the game, we introduce a new equilibrium concept – perfect psychological Bayesian equilibrium, which further extends Dufwenberg and Kirchsteiger (2004) to incomplete information games. Dufwenberg and Kirchsteiger (2004) model, in its turn, is built within the psychological game theory framework (due to Geanakoplos et al. (1989)), which also constitutes the intellectual home of our paper. In psychological game theory, utility is modelled as a function of not only outcomes, but also pre-game beliefs.

The rest of the paper proceeds as follows. Section 2 develops our framework as a psychological game. Section 2.1.1 informally introduces a notion of Perfect Psychological Bayesian Equilibrium. We postpone the rigorous definition including consistency requirements for belief hierarchies to the subsequent version of this paper. Sections 3 and 4 study the performance of VCG mechanism under preferences for reciprocity, for the simultaneous and sequential implementation respectively. Section 5 provides a concluding remark.

2 The Model

2.1 Strategies and Beliefs

Let \( (N, \Theta, H, i(h), \Sigma, u) \) be an extensive-form game of incomplete information, where \( N = \{1, 2, \ldots, n\} \) is the set of players, \( \Theta = (\Theta_i) \) is the set of type vectors, \( H \) is the set of histories with a typical element \( h \in H \); \( i(h) : H \rightarrow N \) is a function assigning to each history a player that moves at that history, \( \Sigma = (\Sigma_i) \) is the set of behavioural strategies and \( u = (u_i) \) the set of utility functions. Player \( i \)'s utility \( u_i : \Sigma \times \Theta_i \times B_i \rightarrow R \) is a function over a strategy profile \( \sigma \in \Sigma \), the player’s own type \( \theta_i \in \Theta_i \) and the player’s \( \beta \)-belief \( \beta_i \in B_i = (\Sigma_j|\Theta_j|)_{-i} \times \Pi_i \left( \Pi_j \Sigma_j|\Theta_j| \right)_{-i} \) (as explained below).

There are two types of beliefs in the game: \( \beta \)-beliefs and \( \mu \)-beliefs. We consider them in turn.

\( \beta \)-beliefs are the pre-play beliefs about what other players are going to choose, and about what other players think their opponents are going to choose (first- and second-order beliefs in (Geanakoplos et al., 1989) framework). In the game of complete informa-
tion, the first-order beliefs are drawn from the strategy space and the second-order beliefs are drawn from the \( n \)-product of strategy space with itself. In the mechanism considered in this note, Player \( i \)'s strategy is a function from the type space into the message space: \( \hat{\theta}_i(\theta_i) : \{\theta_L, \theta_H\} \rightarrow \{\hat{\theta}_L, \hat{\theta}_H\} \). Player \( i \)'s first-order belief specifies what he thinks his opponent of either type will report, hence it is also a function from the type space into the message space: \( \hat{\theta}_j(\theta_j) : \{\theta_L, \theta_H\} \rightarrow \{\hat{\theta}_L, \hat{\theta}_H\} \). Finally, Player \( i \)'s second-order belief specifies what he thinks his opponent of either type believes about his own (player \( i \)'s) report; for both of his types: \( \hat{\theta}_{ii}(\theta_i, \theta_j) : \{\theta_L, \theta_H\} \times \{\theta_L, \theta_H\} \rightarrow \{\hat{\theta}_L, \hat{\theta}_H\} \times \{\hat{\theta}_L, \hat{\theta}_H\} \).

\( \mu \)-beliefs relate to non-singleton information sets. At each information set \( h \in H \), the belief \( \mu \) specifies, for the player \( i(h) \) who moves at that set, his probability assessment of the relative likelihoods of being at each node \( x \in h \) of that set. In our mechanism, the only source of strategic uncertainty are the players’ types. Hence, Player \( i \)'s \( \mu \)-belief at history \( h \) (denoted \( \mu_i(h) \)) specifies his belief that his opponent’s type is high.

There is a connection between \( \beta \)-beliefs and \( \mu \)-beliefs. In an extensive form game, a player’s first-order \( \beta \)-belief induces a probability distribution over decision nodes in each information set reached with positive probability according to this belief. That is, given a first-order \( \beta \)-belief, a player will have a \( \mu \)-belief about his opponents’ type in each information set reached with positive probability. In our game, if a second mover thinks that a high opponent reports high and a low opponent reports low, then, after observing a high report by the first mover, he would believe with certainty that he is of a high type. A difficulty arises in information sets which, according to the first-order belief, are never reached. Suppose, for example, that a second mover believes that the first mover always reports low, irrespective of their type. What would be the second mover’s belief about the first mover’s type, if the first mover reports high? This probability cannot be derived from the belief hierarchy. Our equilibrium concept is agnostic about off-equilibrium beliefs; hence it is called Perfect Psychological Bayesian Equilibrium (PPBE).

### 2.1.1 Perfect Psychological Bayesian Equilibrium

A Perfect Psychological Bayesian Equilibrium of an extensive-form psychological game of incomplete information \( \langle N, \Theta, H, i(h), \Sigma, u \rangle \) is a triple \( (\sigma, \beta, \mu) \) such that

- \( \sigma \) is sequentially rational given \( \mu \) (that is, each player of each type is optimising at each information set, given their belief about their node within this information set captured by \( \mu \));
• $\mu$ is obtained from $\beta$ via Bayes’ rule, whenever possible;
• pre-play beliefs are correct ($\beta$ is obtained from $\sigma$).

2.2 Pivot Mechanism

Consider a standard two-agent setting ($N = \{1, 2\}$) where each agent has either a high, or low valuation of the public good: $\Theta_i = \{\theta_L, \theta_H\}$. The types are independently uniformly distributed: $p_H = p_L = 0.5$.

In the message game, the agents strategically report their types to the direct mechanism. Let $\hat{\theta} = (\hat{\theta}_i, \hat{\theta}_j) \in \Theta_i \times \Theta_j$ be a vector of reported types. The mechanism consists of a public good provision rule ($q(\hat{\theta})$) and a transfer rule ($t(\hat{\theta})$).

Let the cost of public good be $c$. The rule for public good provision is standard for pivot mechanisms. The good is produced ($q = 1$) if and only if the sum of agents’ reported valuations of the good exceeds its cost (1). When the good is not produced, we write $q = 0$.

\[ q = 1 \text{ iff } \hat{\theta}_1 + \hat{\theta}_2 > c. \quad (1) \]

We assume that it is efficient to produce the good when at least one of the agents has a high valuation of the good:

\[ 2\theta_L < c < \theta_L + \theta_H. \quad (2) \]

Given assumption (2), the public good provision rule (1) is summarised in Table 1.

<table>
<thead>
<tr>
<th>$\hat{\theta}_1$</th>
<th>$\theta_L$</th>
<th>$\theta_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_L$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\theta_H$</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Public Good Provision Rule

Assumption (2) lets us concentrate on an interesting case where it is efficient for a high-type agent to provide the good on their own. However, in the presence of reciprocity motives, they may still prefer to report low demand.

According to the transfer rule (3), an agent’s payment does not depend on their 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:

\begin{equation}
    t_i(\hat{\theta}) = \theta_L q(\hat{\theta}_L, \hat{\theta}_{-i}) + \left( q(\hat{\theta}) - q(\hat{\theta}_L, \hat{\theta}_{-i}) \right) (c - \hat{\theta}_{-i}).
\end{equation}

(3)

Material payoff of an agent equals their valuation of the public good (if produced) minus the transfer they have to pay:

\begin{equation}
    \pi_i = \theta_i q(\hat{\theta}) - t_i(\hat{\theta}).
\end{equation}

(4)

2.3 Dufwenberg and Kirchsteiger (2004) Sequential Reciprocity Model

The material payoff described in the previous section is only one part of the agent’s motivation. In addition to a material considerations, agents are also motivated by reciprocity, which we model using a Dufwenberg and Kirchsteiger (2004) model (adapted here for a two-player game):

\begin{equation}
    u_i = \pi_i + Y_i \kappa_i \lambda_i,
\end{equation}

(5)

where \( u_i \) is agent \( i \)’s overall utility, \( \pi_i \) is his material payoff, \( Y_i \in \mathbb{R}^+ \) is \( i \)’s sensitivity to reciprocity (exogenously given), \( \kappa_i \) is \( i \)’s kindness to \( j \) and \( \lambda_i \) is \( i \)’s belief about \( j \)’s kindness to \( i \) (to be explained below).

In what follows, we apply the Dufwenberg and Kirchsteiger (2004) to both a simultaneous- and a sequential-reporting mechanisms. In the first step, we show that simultaneous-reporting VCG is not incentive compatible under utility (5).

3 Simultaneous Reporting

It is a well-known fact that pivot mechanism dominant strategy implements truthful reporting (under selfish preferences). As we show below, it is not incentive compatible when agents are motivated by reciprocity, as modelled by Dufwenberg and Kirchsteiger (2004).

Claim. Consider the pivot mechanism in a two-agent Public Good Provision game with transfer rule (3) and public good provision rule (1). If agents exhibit preference for
reciprocity (as defined in Dufwenberg and Kirchsteiger (2004)), it is not a dominant strategy to report truthfully (for some values of parameters). Specifically, in one equilibrium, the agents report low demand irrespective of their true demand.

**Proof.** In order to prove that truthful reporting is no longer a dominant strategy, we need to fully specify agents’ preferences, including the reciprocity component. We assume agents’ overall utility function is described by model (5), i.e. is a sum of material payoff \( \pi_i \) and the reciprocity component \( Y_i \kappa_i \lambda_i \), where the material payoff equals an agent’s enjoyment from the public good (if produced) minus the transfer he has to pay: \( \pi_i = \theta_i q(\hat{\theta}) - t_i(\hat{\theta}) \). Thus, \( \pi_i \) is the function of agent \( i \)'s true type and the reported type profile.

The transfer rule for the two-player, two-type pivot mechanism can be expanded as follows:

\[
\begin{align*}
t_1(\hat{\theta}_L, \hat{\theta}_L) &= t_2(\hat{\theta}_L, \hat{\theta}_L) = \theta_L \cdot 0 + (0 - 0)(c - \theta_L) = 0; \\
t_1(\hat{\theta}_L, \hat{\theta}_H) &= t_2(\hat{\theta}_H, \hat{\theta}_L) = \theta_L \cdot 1 + (1 - 1)(c - \theta_L) = \theta_L; \\
t_1(\hat{\theta}_H, \hat{\theta}_L) &= t_2(\hat{\theta}_L, \hat{\theta}_H) = \theta_L \cdot 0 + (1 - 0)(c - \theta_L) = c - \theta_L; \\
t_1(\hat{\theta}_H, \hat{\theta}_H) &= t_1(\hat{\theta}_H, \hat{\theta}_H) = \theta_L \cdot 1 + (1 - 1)(c - \theta_H) = \theta_L.
\end{align*}
\] (6)

The reciprocity component of an agent’s utility is a product of his opponent’s kindness to him, and his kindness to the opponent. This component is maximised when an agent’s kindness has the same sign as the opponent’s kindness, and the maximum magnitude. In words, an agent would like to be kind to a kind opponent, and mean to a mean opponent.

The opponent’s kindness to player \( i \) is a function of agent \( i \)'s first-order belief \( \hat{\theta}_j^\prime \) and second-order belief \( \hat{\theta}_i^\prime \). It is the difference between player \( i \)'s actual payoff (when he chooses \( \hat{\theta}_i^\prime \) and his opponent chooses \( \hat{\theta}_j^\prime \)) and his equitable payoff (when choosing \( \hat{\theta}_i^\prime \)):

\[
\lambda_i(\hat{\theta}_i^\prime, \hat{\theta}_j^\prime) = \pi_i(\hat{\theta}_i^\prime, \hat{\theta}_j^\prime) - \pi_i^e(\hat{\theta}_i^\prime),
\] (7)

where \( i \)'s equitable payoff is calculated as follows:

\[
\pi_i^e(\hat{\theta}_i^\prime) = \frac{1}{2} \cdot \left[ \max \left\{ \pi_i(\hat{\theta}_j^\prime, \hat{\theta}_i^\prime) \mid \hat{\theta}_j^\prime \in \Sigma_j^E \right\} + \min \left\{ \pi_i(\hat{\theta}_j^\prime, \hat{\theta}_i^\prime) \mid \hat{\theta}_j^\prime \in \Sigma_j^E \right\} \right].
\] (8)
In words, i’s equitable payoff is equal to the average between the highest and the lowest payoffs i can get, given his belief about j’s belief about his report \( \hat{\theta}_i \). The maxima and minima are taken with respect to j’s efficient strategies \( \Sigma^E_j \). We use Dufwenberg and Kirchsteiger (2004) definition of an efficient strategy as a strategy such that there does not exist some other strategy which, for any strategies of the opponents, brings every player at least as large a payoff, and a strictly larger payoff for at least one player.

The player’s own kindness to his opponent is defined analogously:

\[
\kappa_i(\hat{\theta}_i, \hat{\theta}'_j) = \pi_j(\hat{\theta}_i, \hat{\theta}'_j) - \pi^*_i(\hat{\theta}_i),
\]

where i’s equitable payoff is defined similarly to (8).

We need to make small additions to Dufwenberg and Kirchsteiger (2004) framework in order to account for incomplete information. When player i is assessing his opponent’s kindness, he believes that the opponent does not know i’s true type. Hence, he estimates the opponent’s kindness as an average kindness:

\[
\lambda_i = \mu_j \lambda_i(\theta_H) + (1 - \mu_j) \lambda_i(\theta_L),
\]

where \( \lambda_i(\theta_i) \) is j’s kindness to i when i’s true type is \( \theta_i \) and \( \mu_j \) is i’s estimation of j’s belief that i is high. In equilibrium, this belief is equal to the common prior \( \mu = 0.5 \).

When an agent is picking his own action, he is assumed to be maximising his expected utility, where the expectation is taken with respect to the opponent’s type:

\[
E(u_i) = \pi_i + Y_i \left[ \mu \kappa^h_i \bar{\lambda}^h_i + (1 - \mu) \kappa^l_i \bar{\lambda}^l_i \right],
\]

where \( \mu \) is the probability that the opponent is high, \( \kappa^h_i \) (\( \kappa^l_i \)) is player i’s kindness to a high (low) opponent, and \( \bar{\lambda}^h_i \) (\( \bar{\lambda}^l_i \)) is an average kindness of a high (low) opponent.

Let us check if agents are incentivised to report low when their true type is high. In order to do so, we will determine whether the following type-contingent strategy profile can be a part of an equilibrium: both agents report low irrespective of their true type (12).

\[
\hat{\theta}_i(\theta_L) = \hat{\theta}_i(\theta_H) = \hat{\theta}_L.
\]

In psychological games, an equilibrium also includes a belief profile. Existing psycho-
logical games literature does not provide an equilibrium notion suitable for incomplete information framework, whether in normal or extensive-form games. However, complete-information Psychological Nash Equilibrium can be naturally extended to incomplete information games by defining players’ beliefs as functions of their opponent’s type, as is done in Section 2.1.1 where the notion of Perfect Psychological Bayesian Equilibrium is introduced. Recall that in a psychological game of incomplete information a player’s first-order belief should specify what he thinks his opponent of each type does, and his second-order belief should specify what he thinks his opponent of either type thinks he does, in any possible type.

In equilibrium agents’ beliefs should correspond to reality. Thus, agent $i$’s first-order belief should be $\hat{\theta}_i'(\theta_L) = \hat{\theta}_i'(\theta_H) = \hat{\theta}_L$ (he believes his opponent reports low irrespective of his type), and second-order belief should also be $\hat{\theta}_i''(\theta_i, \theta_j) = \hat{\theta}_L$ for all $i, j$ (he believes his opponent believes he reports low irrespective of his type). In the presence of incomplete information, each agent treats other’s true type probabilistically.

As a part of candidate PPBE described above, let us determine agent 1’s best response to such a strategy and belief profile when he is himself of high type.

In order to compare utilities from reporting high ($\hat{\theta}_1 = \hat{\theta}_H$) VS reporting low ($\hat{\theta}_1 = \hat{\theta}_L$), we need to calculate kindness functions.

**Agent 1’s belief about agent 2’s kindness** depends on his first- and second-order beliefs only, but not on his actual strategy.

If 2 thinks 1’s true type is high, his kindness to him by reporting low is the difference between 1’s actual payoff as a result of such report and 1’s equitable payoff:

$$\lambda_1(\hat{\theta}_L, \hat{\theta}_L; \theta_H) = \pi_1(\hat{\theta}_L, \hat{\theta}_L) - \pi_1^e(\hat{\theta}_L) = 0 - \frac{1}{2} \cdot (\theta_H - \theta_L) = -\frac{\theta_H - \theta_L}{2},$$

where 1’s equitable payoff $\pi_1^e(\hat{\theta}_L)$ equals the average of his minimum payoff of 0 (if 2 reports low) and his maximum payoff of $\theta_H - \theta_L$ (if 2 reports high).

The calculation of the equitable payoff uses the fact that both strategies of a low player 2 (reporting high and low) are efficient: if he reports high, he will give a high-reporting player 1 a higher payoff. If he reports low, he will give himself a higher payoff if 1 reports low. Both strategies of a high player 2 (reporting high and low) are also efficient: if he reports high, he will give a high-reporting player 1 a higher payoff. If he reports low, he will give himself a higher payoff in case 1 reports high. By the same logic, both strategies
of 1 of either type are efficient.

If 2 thinks 1’s true type is low, his kindness to him by reporting low is the difference between 1’s actual payoff as a result of such report (0) and 1’s equitable payoff:

\[ \lambda_1(\hat{\theta}_L, \hat{\theta}_L; \theta_L) = \pi_1(\hat{\theta}_L, \hat{\theta}_L; \theta_L) - \pi_1^e(\hat{\theta}_L; \theta_L) = 0 - 0 = 0, \]  

(14)

where 1’s equitable payoff is zero because a low, low-reporting 1 is receiving a payoff of 0 regardless of 2’s payoff.

Player 2’s average kindness from reporting low is thus

\[ \bar{\lambda}_1 = \frac{1}{2} \left( \lambda_1(\hat{\theta}_L, \hat{\theta}_L; \theta_H) + \lambda_1(\hat{\theta}_L, \hat{\theta}_L; \theta_H) \right) = \frac{1}{2} \left( -\frac{\theta_H - \theta_L}{2} + 0 \right) = -\frac{\theta_H - \theta_L}{4}. \]  

(15)

**Agent 1’s kindness to 2** depends on 1’s strategy \( \hat{\theta}_1 \), 1’s belief about 2’s strategy \( \hat{\theta}_2 \) and 2’s type \( \theta_2 \).

- If 2 is high, 1’s kindness is as follows:
  - If 1 reports low: \( \kappa_1^H = 0 - \frac{1}{2} (\theta_H - \theta_L) = -\frac{1}{2} (\theta_H - \theta_L) \).
  - If 1 reports high: \( \kappa_1^H = (\theta_H - \theta_L) - \frac{1}{2} (\theta_H - \theta_L) = \frac{1}{2} (\theta_H - \theta_L) \).

- If 2 is low, 1’s kindness is as follows:
  - If 1 reports low: \( \kappa_1^L = 0 - 0 = 0 \).
  - If 1 reports high: \( \kappa_1^L = 0 - 0 = 0 \).

We are now ready to determine which action (reporting high or low) yields bigger expected overall utility for 1. Materially, 1 is better off reporting his true type (high), as proved by incentive compatibility of the pivot mechanism. When reciprocity is considered, his expected overall utility from reporting is

\[
\begin{align*}
    \mathbb{E}_{\theta_2} \left( u_1(\hat{\theta}_H, \hat{\theta}_L) \right) &= \theta_H - (c - \theta_L) + Y_1 \left( -\frac{\theta_H - \theta_L}{4} \right) \left[ 0.5 \cdot \frac{\theta_H - \theta_L}{2} + 0.5 \cdot 0 \right]; \\
    \mathbb{E}_{\theta_2} \left( u_1(\hat{\theta}_L, \hat{\theta}_L) \right) &= 0 + Y_1 \left( -\frac{\theta_H - \theta_L}{4} \right) \left[ 0.5 \cdot \left( -\frac{\theta_H - \theta_L}{2} \right) + 0.5 \cdot 0 \right],
\end{align*}
\]

(16)

which yields
\[
E_{\theta_2} \left( u_1(\hat{\theta}_H, \hat{\theta}_L) \right) = \theta_H - \theta_L - c - \frac{Y_i(\theta_H - \theta_L)^2}{16};
\]
\[
E_{\theta_2} \left( u_1(\hat{\theta}_L, \hat{\theta}_L) \right) = \frac{Y_i(\theta_H - \theta_L)^2}{16}.
\]

(17)

Reporting low is a best response when \( u_1(\hat{\theta}_H, \hat{\theta}_L) > u_1(\hat{\theta}_L, \hat{\theta}_L) \), that is, when the following condition holds:

\[
Y_i \geq 8 \frac{\theta_H - \theta_L - c}{(\theta_H - \theta_L)^2}.
\]

(18)

Thus, if agent 1’s reciprocity sensitivity is strong enough, he will be willing to misreport. As the game is symmetric, the same is true of agent 2. Subsequently, reporting low when being of high type can be part of an equilibrium of the pivot mechanism under large enough reciprocity sensitivity. \textbf{Q.E.D.}

We can already see that in the above example the mechanism does not fully implement (neither in dominant strategies, nor in PPBE) truthful reporting. This is because, in the case both players are of high type, they are incentivised to misreport if they believe their opponent is reporting low. It is easy to show that reporting low is each player’s best response when their true type is low. Hence, both players reporting low irrespective of their true type is an equilibrium in the pivot mechanism with reciprocity.

It is also straightforward to show that truthful reporting is also an equilibrium (under all values of parameters). Indeed, when your opponent is reporting his true type, he is being neither generous nor unkind to you, \textit{i.e.} he is neither sacrificing anything to help you nor free-rides on you. Thus, you evaluate his kindness to you as zero. This renders the whole reciprocity component of your overall utility zero. Thus, your objective is to maximise the material component of utility function, which can be achieved by truthful reporting under the pivot mechanism.

To sum up, when agents exhibit preferences for reciprocity, the pivot mechanism induces either one, or two Perfect Psychological Bayesian Equilibria: truthful reporting (irrespective of values of parameters) and under-reporting (under some values of parameters).
4 Sequential Reporting

Can reporting low irrespective of one’s type still be an equilibrium under sequential reporting, when the first mover’s report announcement becomes public information before the second mover makes their choice? In what follows, we show that sequentality help get rid of the adverse under-reporting equilibrium. Reciprocity holds the key to this positive result: if the first mover deviates by reporting high, the second mover would feel grateful and would be willing to reciprocate, also reporting high and bringing both players higher payoffs than in the under-reporting equilibrium.

Fix a strategy and a consistent belief profile where both agents always report low, and consider a deviation by the first mover, i.e. suppose he reports high. How would this action be judged by the second mover?

The second mover assesses a high report by the first mover as kind:

\[
\bar{\lambda}_2 = \frac{1}{2} \left( \lambda_2(\hat{\theta}_L, \hat{\theta}_L; \theta_H) + \lambda_2(\hat{\theta}_L, \hat{\theta}_L; \theta_L) \right) = \frac{1}{2} \left( \frac{\theta_H - \theta_L}{2} + 0 \right) = \frac{\theta_H - \theta_L}{4}, \quad (19)
\]

where \( \lambda_1(\hat{\theta}_L, \hat{\theta}_L; \theta_H) \) is the first mover’s kindness to the high second mover, \( \lambda_1(\hat{\theta}_L, \hat{\theta}_L; \theta_L) \) is the first mover’s kindness to the low second mover and these values are calculated as follows:

\[
\lambda_2(\hat{\theta}_L, \hat{\theta}_L; \theta_H) = \theta_H - \theta_L - \frac{(\theta_H - \theta_L) + 0}{2} = \frac{\theta_H - \theta_L}{2}, \quad (20)
\]

\[
\lambda_2(\hat{\theta}_L, \hat{\theta}_L; \theta_L) = 0 - \frac{0 + 0}{2} = 0. \quad (21)
\]

Note that neither strategy of the first mover is inefficient; hence the second mover’s equitable payoff is equal to the average of his payoffs when the first mover chooses high or low.

Let us now determine the second mover’s kindness (remember we are looking at the high second movers only). Her kindness to the first mover depends on the second mover’s action, and the first mover’s type (believed to be high with probability \( \mu \)):

- If 1 is high, 2’s kindness is as follows:
  - If 2 reports low: \( \kappa_2^H = \theta_H - (c - \theta_L) - (\theta_H - \theta_L) = 2\theta_L - c \).
  - If 2 reports high: \( \kappa_2^H = (\theta_H - \theta_L) - (\theta_H - \theta_L) = 0 \).
• If 1 is low, 2’s kindness is as follows:

- If 2 reports low: \( \kappa^L_2 = (\theta_L - (c - \theta_L)) - (\theta_L - \theta_L) = 2\theta_L - c. \)
- If 2 reports high: \( \kappa^H_2 = (\theta_L - \theta_L) - (\theta_L - \theta_L) = 0. \)

Note that reporting low after the first mover reports high is an inefficient strategy for a high second mover, as reporting high in that decision node would have improved the first mover’s payoff and left the second mover’s payoff unchanged.

We are now ready to determine which action (reporting high or low) yields bigger expected overall utility for the high second mover. Materially, 2 is indifferent between reporting high or low, as his report is not going to be pivotal. However, when reciprocity is considered, his expected overall utility from reporting is

\[
\begin{align*}
E_{\theta_2}(u_2(\hat{\theta}_H, \hat{\theta}_L)) &= \theta_H - \theta_L + Y_2^\prime \frac{(\theta_H - \theta_L)(2\theta_L - c)}{4}; \\
E_{\theta_2}(u_2(\hat{\theta}_H, \hat{\theta}_L)) &= \theta_H - \theta_L + Y_2^\prime \frac{\theta_H - \theta_L}{4} \cdot 0 = \theta_H - \theta_L.
\end{align*}
\]

Hence, reporting high is always the best response for the second mover if she cares about reciprocity \((Y_2 > 0)\). Knowing this, a high first mover will deviate and report high, even if he is completely selfish \((Y_1 = 0)\). Hence, always reporting low cannot be an equilibrium of the game when the second mover is motivated by reciprocity as modelled by Dufwenberg and Kirchsteiger (2004).

5 Concluding Remark

This paper showed that sequential implementation of a 2-person pivot mechanism when agents are motivated by reciprocity (Dufwenberg and Kirchsteiger, 2004) eliminates an inefficient equilibrium present in the simultaneous case, where both agents underreport their true demand. Hence, the sequential mechanism with reciprocity performs better than both the simultaneous mechanism with reciprocity and the sequential mechanism with selfish agents.

References


