Journal of Economic Theory (2000) 91, 180-198.

Manipulation through Bribes

James Schummer*

MEDS Department, Kellogg Graduate School of Management

Northwestern University, Evanston, IL 60208–2009

First version: June 1997

Revised: April 2, 1999

Abstract

We consider allocation rules that choose both an outcome and transfers, based on the agents' reported valuations of the outcomes. Under a given allocation rule, a bribing situation exists when agent j could pay agent i to misreport his valuations, resulting in a net gain to both agents. A rule is *bribe-proof* if such opportunities never arise.

The central result is that when a *bribe-proof* rule is used, the resulting payoff to any one agent is a *continuous* function of any other agent's reported valuations. We then show that on connected domains of valuation functions, if either the set of outcomes is finite or each agent's set of admissible valuations is *smoothly* connected, then an agent's payoff is a *constant* function of other agents' reported valuations. Finally, under the additional assumption of a standard domain-richness condition, we show that a *bribe-proof* rule must be a constant function. The results apply to a very broad class of economies.

Keywords: Strategy-proof, Bribe, Manipulation, Collusion.

JEL Classification Numbers: C70, D70.

*Discussion with John Duggan was helpful. I also thank Matt Jackson, Shigehiro Serizawa, members of the Math Center Bag Lunch Seminar at Northwestern, and anonymous referees for comments.

1 Introduction

Consider the problem of choosing an outcome (e.g., a level of public goods, an allocation of private goods, an assignment of jobs, etc.) and transfers of a private good among agents, based on the way the involved agents value the outcomes (or on other characteristics). Without knowing the agents' valuations, it may not be reasonable for the decision maker to assume that they would truthfully reveal them if doing so is not in their best interest. For example, this is a concern in the literature on *strategy-proof* allocation rules for public goods economies (Clarke, 1971; Groves, 1973; Green and Laffont, 1977; Moulin, 1980), exchange economies (Hurwicz, 1972; Barberà and Jackson, 1995), and many other domains. Most of this work addresses this incentives problem at the individual level in combination with standard distributional requirements (e.g., efficiency). Our concern here is to address a *group* incentives problem in the absence of any distributional requirements.

One compelling group incentive-compatibility requirement is *coalitional strategy-proofness*: no coalition of agents should be able to jointly misrepresent their valuations (or other characteristics) in a way that results in a direct gain to each of those agents. The desirability of this condition is clear. In many environments, however, this condition is too strong, ruling out all but a few, unreasonable decision rules.[2]

One way that *coalitional strategy-proofness* is too strong is that even very large coalitions are prevented from manipulating. The execution of a joint misrepresentation by a large number of agents requires not only that they know each others' valuations, but also that they are able to coordinate their actions. In many situations, it is not practical to worry about such types of manipulation.

On the other hand, if we suspect that a *small* coalition could manage to coordinate their actions in a profitable, joint misrepresentation of valuations, then they could manage to, additionally, arrange transfers to each other, i.e. they could bribe each other to misrepresent.

To address these two concerns, we formulate the weakest intuitive condition that rules out this type of misrepresentation by coalitions of size two (or one): Only one agent is bribed by one other agent to solely misrepresent

[2] Consider, for example, the well-established coalitional manipulability of Clarke–Groves mechanisms.

his valuations. There are only two agents involved in the transfer, and only one of them misrepresents his type. Decision rules that eliminate the possibility of this type of manipulation—and that are *strategy-proof*—are called *bribe-proof*.[3]

The results apply generally to settings in which the outcomes may represent levels of public goods, allocations of private goods, matching assignments, etc., as long as the agents have a transferable, divisible good with which to bribe each other, and they have continuous preferences that are additive in the transferable good.

The central result is that in a very general setting, if an allocation rule is *bribe-proof*, then it satisfies a continuity property: The payoff that an agent receives varies continuously with respect to changes in the reported valuations of any other agent.

From this we show that if the set of outcomes is finite, then a *bribe-proof* allocation rule is "essentially" constant, in the sense that an agent's payoff is never affected by a change in any other agent's reported valuations. This result applies not only when all possible valuations of the outcomes are admissible, but also when each agent's set of admissible valuations is a connected set. For the case of an infinite set of outcomes, we derive the same conclusion as long as each agent's set of admissible valuations is "smoothly connected" in the sense of Holmström (1979), who generalizes the classic characterization of Clarke–Groves mechanisms to such domains.

Finally, we show that if the domain of valuation functions is sufficiently "rich", then a *bribe-proof* allocation rule must actually be a constant function. In the Conclusion, we discuss the relation between these results and those of Crémer (1996), who works explicitly in an imperfect information environment. We also discuss broader possible interpretations of our results.

For the case of exactly two public good outcomes (and the unrestricted domain of valuation functions), Green and Laffont (1979) consider manipulations by coalitions of a fixed size in which members make *joint* misrepresentations, along with transfers among themselves. They show that no Clarke–Groves mechanism is immune to such manipulation by coalitions of

[3] It may be more natural to attempt to rule out situations in which two agents *jointly* misrepresent. Since our results concern the effects of imposing *bribe-proofness*, they are technically stronger by using the weaker version of the condition. Serizawa (1998) examines the consequences of disallowing pairs of agents to jointly misrepresent types with no transfers.

any fixed size less than the total number of agents. Our result for finite sets of outcomes and connected domains of valuations is a substantial strengthening of their result for the particular case of manipulation by coalitions with a size of two.

2 Model

There is a finite set of **agents**, $N = \{1, 2, \ldots, n\}$, $n \geq 2$, with arbitrary elements i and j. There is a compact set of **outcomes**, $Y \subseteq \mathbb{R}^{\ell}$, with arbitrary elements y and y'. In addition to the outcome, each agent $i \in N$ consumes some amount $m_i \in \mathbb{R}$ of a divisible good, say money. An **allocation** consists of an outcome $y \in Y$ and the specification of an amount of money for each agent, $m = (m_1, m_2, \ldots, m_n)$.

Each agent $i \in N$ has a quasi-linear (i.e. linear additively separable) preference ordering over $Y \times \mathbb{R}$. We will assume that these preferences can be parameterized in a continuous way by a set, $\Theta_i \subseteq \mathbb{R}^k$, of admissible **types**. That is, for each $i \in N$, there is a continuous **valuation function**, $v_i \colon Y \times \Theta_i \to \mathbb{R}$, representing the preferences of an agent, depending on his type. An agent of type $\theta_i \in \Theta_i$ (weakly) prefers a bundle $(y, m_i) \in Y \times \mathbb{R}$ to another bundle (y', m_i') if and only if $v_i(y; \theta_i) + m_i \geq v_i(y'; \theta_i) + m_i'$.

When Y is finite, one could imagine the set of types being $\mathbb{R}^{|Y|-1}$, and that a type represents an agent's normalized valuations of the public outcomes. More generally, the set may also describe other characteristics of an agent. When modeling a certain class of economies, such a parameterization may not be unique. All that is required for our results to hold is that such a continuous parameterization *exists*. The assumption that Y and Θ_i are subsets of Euclidean spaces was made to avoid the distraction of topological generality.

A **domain** is a cross-product of type sets $\Theta \equiv \Theta_1 \times \cdots \times \Theta_n$. For any $i \in N$, we use the standard notation $\theta_{-i} \in \Theta_{-i}$ to refer to a list of types for agents other than i.

Some of the results apply to domains that are connected: The domain Θ is **connected** if for all $i \in N$, Θ_i is a path-connected set, i.e., for all $\theta_i, \theta_i' \in \Theta_i$, there exists a continuous function $f \colon [0, 1] \to \Theta_i$ such that $f(0) = \theta_i$ and $f(1) = \theta_i'$.

Our general allocation problem then is a specification of N, Y, Θ, and

$\{v_i\}_{i\in N}$. A **solution** (to this problem) is a function $\varphi\colon \Theta \to Y \times \mathbb{R}^n$, choosing an allocation for any profile of admissible types. It will be convenient to decompose the solution into two functions, $\bar{y}\colon \Theta \to Y$ and $\bar{m}\colon \Theta \to \mathbb{R}^n$, in which case we write $\varphi \equiv (\bar{y}, \bar{m})$. It will also be convenient to write, for any agent $i \in N$, $\varphi_i(\theta) \equiv (\bar{y}(\theta), \bar{m}_i(\theta))$ to refer to agent i's consumption bundle.

Depending on the interpretation of the model, one may wish to additionally impose certain feasibility conditions on a solution, such as *weak budget balance* (for all $\theta \in \Theta$, $\sum \bar{m}_i(\theta) \leq M$, where M is some aggregate endowment of money), or *strong budget balance* ($\sum \bar{m}_i(\theta) = M$). Such requirements have no effect on our results, so we will not address them.

2.1 The Bribing Condition

If the agents are reporting their types to a planner who is using a given solution, it is of interest to know whether the solution satisfies certain incentive compatibility properties. For instance, it is desirable for a solution to be such that an agent of type θ_i can do no better for himself than by reporting his true type θ_i to the planner, regardless of the other agents' types. That is, a solution should satisfy the following condition.

Strategy-proof: The solution $\varphi = (\bar{y}, \bar{m})$ is *strategy-proof* if for all $\theta \in \Theta$ and all $i \in N$, there exists no $\theta_i' \in \Theta_i$ such that

$$v_i(\bar{y}(\theta_i', \theta_{-i}); \theta_i) + \bar{m}_i(\theta_i', \theta_{-i}) \ > \ v_i(\bar{y}(\theta); \theta_i) + \bar{m}_i(\theta)$$

Holmström (1979) shows that on most connected domains, the only *strategy-proof* solutions that maximize $\sum_i v_i(y; \theta_i)$ for every profile θ are Clarke–Groves mechanisms.

As discussed in the Introduction, we also wish to rule out the possibility that an agent could bribe another to misrepresent his type. We formulate a condition that rules out this type of situation.

Bribe-proof: The solution $\varphi = (\bar{y}, \bar{m})$ is *bribe-proof* if for all $\theta \in \Theta$ and all $i, j \in N$, there exists no $b \in \mathbb{R}$ and $\theta_i' \in \Theta_i$ such that

$$v_i(\bar{y}(\theta_i', \theta_{-i}); \theta_i) + \bar{m}_i(\theta_i', \theta_{-i}) + b \ > \ v_i(\bar{y}(\theta); \theta_i) + \bar{m}_i(\theta)$$
$$v_j(\bar{y}(\theta_i', \theta_{-i}); \theta_j) + \bar{m}_j(\theta_i', \theta_{-i}) - b \ > \ v_j(\bar{y}(\theta); \theta_j) + \bar{m}_j(\theta)$$

The interpretation is that j bribes i with b units of money to misrepresent his type. Notice that by choosing $i = j$ and $b = 0$, *bribe-proofness* implies *strategy-proofness*.[4]

At this point, a few points are worth mentioning. First, we are implicitly assuming that the two agents would trust each other in arranging this misrepresentation; that is, j would not break his promise to pay i, and i would not renege on his promise to misrepresent. Therefore, the results of this paper should be seen as a comment on the planner's "worst-case scenario" in which agents completely trust each other. See Tirole (1992) for a discussion of the assumption. We will discuss a departure from this assumption in the Conclusion.

Second, note that we do not allow j to also misrepresent his type. Since we are determining the consequences of our condition (which in some instances are strong), the results are stronger using our weaker definition.

Third, we are implicitly assuming that the divisible good is perfectly transferable among agents. Perhaps instead, when j sends b units of money, i only receives $\lambda b < b$ of it (e.g. due to some transaction cost). All of our results continue to hold as long as $\lambda > 0$.

3 Results

Consider the following example of a domain for which there exist non-trivial *bribe-proof* solutions.

Example 1 (Symmetric, single-peaked preferences.) Let $Y = [a, b] \subset \mathbb{R}$ be an interval of public outcomes. For all $i \in N$, let $\Theta_i = \mathbb{R}$, and for all $y \in Y$ and all $\theta_i \in \Theta_i$, let $v_i(y, \theta_i) = -|y - \theta_i|$. Note that each $v_i(\,\cdot\,, \theta_i)$ is a single-peaked function on Y and that Θ is a connected domain.

A *bribe-proof* solution $\varphi = (\bar{y}, \bar{m})$ can be constructed by letting $\bar{m}$ be constant and letting $\bar{y}$ be defined as a median voter rule (Moulin, 1980), e.g., if $|N|$ is odd, let $\bar{y}(\theta)$ be the median of $\theta_1, \ldots, \theta_n$.

[4] One might argue that the *bribe-proofness* condition should be defined without implying *strategy-proofness* (i.e. disallowing $i = j$). It would be unreasonable, however, to attempt to rule out bribing situations while allowing the possibility that an agent could gain by simply misrepresenting his own type.

We will point out particular attributes of this example in Sections 3.2–3.3 that have allowed us to construct non-trivial *bribe-proof* solutions. First, however, note that under any of these solutions, an agent's payoff varies continuously as any other agent varies his reported type. We first show that this is a general property of *bribe-proof* solutions.

3.1 Continuity

Our analysis concerns the effect of a change in the report of an agent's type on a given *bribe-proof* solution. Therefore, **for the remainder of the paper, we fix the following**.

- an agent $i \in N$ (potentially the bribee),
- types $\theta_{-i} \in \Theta_{-i}$ of the other agents,
- an agent $j \in N \setminus \{i\}$ (potentially the briber), and
- a *bribe-proof* solution $\varphi \equiv (\bar{y}, \bar{m})$.

The set of bundles that agent i can obtain by varying his type is his **option set**:

$$\bar{O}_i = \{(y, m_i) \in Y \times \mathbb{R} : \exists \theta_i \in \Theta_i \text{ such that } (y, m_i) = \varphi_i(\theta_i, \theta_{-i})\}$$

With respect to any $\theta_i \in \Theta_i$, define the maximum payoff that i may receive, and his set of best obtainable bundles, as follows.

$$\begin{aligned} u_i^*(\theta_i) &= \max_{(y,m_i)\in\bar{O}_i} v_i(y;\theta_i) + m_i \\ O_i^*(\theta_i) &= \{(y, m_i) \in \bar{O}_i : v_i(y;\theta_i) + m_i = u_i^*(\theta_i)\} \end{aligned}$$

Since φ is *strategy-proof*, u_i^* is well-defined; in fact, for all $\theta_i \in \Theta_i$ we have

$$u_i^*(\theta_i) = v_i(\bar{y}(\theta_i, \theta_{-i}); \theta_i) + \bar{m}_i(\theta_i, \theta_{-i})$$

We first show that since φ is *strategy-proof*, u_i^* must be a continuous function.[5] If Y is finite, this follows from a direct application of the Maxi-

[5]This general notion that *strategy-proofness* implies some sort of continuity has been accepted by some as "folk knowledge"; however it usually requires an assumption on

mum Theorem. In the general case, however, the option set $\bar{O}_i$ may not be compact, and the Maximum Theorem can not be directly applied.

The following lemma shows that on any compact set of types, an agent's option set—in particular, the amount of the divisible good that he receives—is bounded.

LEMMA 1 *Suppose that* Y *is compact,* Θ_i *is compact, and* φ *is* strategy-proof. *Then* $\bar{m}_i(\,\cdot\,,\theta_{-i})$ *is a bounded function of* Θ_i*.*

Proof: Let $v = \sup_{\theta_i} v_i(\bar{y}(\theta_i,\theta_{-i});\theta_i)$ and $v' = \inf_{\theta_i} v_i(\bar{y}(\theta_i,\theta_{-i});\theta_i)$. Since Y and Θ_i are compact, $v < \inf$ and $v' > -\inf$.

Furthermore, for all $\theta_i, \theta_i' \in \Theta_i$, *strategy-proofness* implies

$$v - v' \geq v_i(\bar{y}(\theta_i,\theta_{-i});\theta_i) - v_i(\bar{y}(\theta_i',\theta_{-i});\theta_i) \geq \bar{m}_i(theta_i',\theta_{-i}) - \bar{m}_i(theta_i,\theta_{-i})$$

Hence $\bar{m}_i(\,\cdot\,,\theta_{-i})$ is bounded. □

Therefore, agent i's option set on a compact domain is a bounded set. It may not, however, be closed. To deal with that technical difficulty, define the following analogs of u_i^* and O_i^*.[6]

$$\begin{aligned} u^{**}(\theta_i) &= \max_{(y,m_i)\in \mathrm{cl}(\bar{O}_i)} v_i(y;\theta_i) + m_i \\ O^{**}(\theta_i) &= \{(y,m_i) \in \mathrm{cl}(\bar{O}_i) : v_i(y;\theta_i) + m_i = u^{**}(\theta_i)\} \end{aligned}$$

PROPOSITION 1 *Suppose that* Y *is compact,* Θ_i *is compact, and* φ *is* strategy-proof. *Then* u_i^* *is continuous in* Θ_i*.*

Proof: Lemma 1 implies that $\mathrm{cl}(\bar{O}_i)$ is compact. Therefore the Maximum Theorem implies that u^{**} is continuous and O^{**} is upper semi-continuous (u.s.c.).[7] Note that by the definition of *strategy-proofness*, u_i^* is in fact well-defined. Therefore, since v_i is continuous on $\mathrm{cl}(\bar{O}_i)$, we have $u_i^* \equiv u^{**}$. Hence u_i^* is continuous. □

the consumption space, such as our compactness assumption on Y; see Example 2. See Chichilnisky and Heal (1997) for a solution-continuity result on single-peaked preferences.

[6]The closure of a set S is denoted $\mathrm{cl}(S)$.

[7]See Berge (1963) for a definition of u.s.c.. The notation follows Sundaram (1996).

It should be noted that this proposition is more of a technical result than an applied one. The function u_i^* measures the payoff to agent i from reporting type θ_i when his true type is θ_i. The proposition states that when this report and true type are varied *together*, the resulting payoff is continuous. The proposition does not imply general continuity resulting from varying a report with respect to a *fixed* true type.

Secondly, note that since u_i^* is continuous on any compact (sub)domain, it is continuous on any domain (see the proof of Corollary 1).

For all $\theta_i \in \Theta_i$, define the payoff that j receives as follows.

$$u_j^*(\theta_i) \;=\; v_j(\bar{y}(\theta_i, \theta_{-i}); \theta_j) + \bar{m}_j(\theta_i, \theta_{-i})$$

The central result is that u_j^* is a continuous function of agent i's reported type. One way to demonstrate this fact is as follows.[8] *Bribe-proofness* requires agent i to maximize the sum $u_i^* + u_j^*$ through his reported type. Viewing this sum as a pseudo-payoff for agent i, *bribe-proofness* implies a *strategy-proofness* with respect to this pseudo-payoff. Hence it follows from Proposition 1 that this sum is continuous, and hence so is u_j^*. Below we provide an alternate formal proof.

THEOREM 1 *Suppose that Y is compact, Θ_i is compact, and φ is* bribe-proof. *Then u_j^* is continuous in Θ_i.*

Proof: Suppose by contradiction that u_j^* is not continuous at some $\bar{\theta}_i \in \Theta_i$.

Case 1: There exists a sequence $\{\theta_i^k\}_{k=1}^{\infty}$ converging to $\bar{\theta}_i$, and $\bar{\epsilon} > 0$, such that for all k, $u_j^*(\bar{\theta}_i) - u_j^*(\theta_i^k) > \bar{\epsilon}$.

Let $(y, m_i) = \varphi_i(\bar{\theta}_i, \theta_{-i})$. *Bribe-proofness* implies that for all k, $u_i^*(\theta_i^k) - v_i(y; \theta^k) - m_i \geq \bar{\epsilon}$. Proposition 1 implies, however, that $u_i^*(\theta_i^k)$ converges to $u_i^*(\bar{\theta}_i)$, and the continuity of v implies that $v_i(y; \theta^k)$ converges to $v_i(y; \bar{\theta}_i)$. Hence $u_i^*(\theta_i^k) - v_i(y; \theta^k) - m_i$ converges to 0, which is a contradiction.

Case 2: There exists a sequence $\{\theta_i^k\}_{k=1}^{\infty}$ converging to $\bar{\theta}_i$, and $\bar{\epsilon} > 0$, such that for all k, $u_j^*(\theta_i^k) - u_j^*(\bar{\theta}_i) > \bar{\epsilon}$.

Since, as above, O^{**} is u.s.c. and $u_i^* \equiv u^{**}$, we have for all $\theta_i \in \Theta_i$, $O_i^*(\theta_i) \subseteq O^{**}(\theta_i)$.

[8] I thank an anonymous referee for pointing out this argument.

Since v_i is continuous, there exists an open set $\hat{O} \supset O^{**}(\bar{\theta}_i)$ such that $(y, m_i) \in \hat{O}$ implies $u_i^*(\bar{\theta}_i) - v_i(y; \bar{\theta}_i) - m_i < \bar{\epsilon}$. Since O_i^{**} is u.s.c., there exists k (sufficiently large) such that $O^{**}(\theta_i^k) \subset \hat{O}$. This implies $(\bar{y}(\theta_i^k), \bar{m}_i(\theta_i^k)) \in \hat{O}$. Therefore, $v_i(\bar{y}(\theta_i^k); \bar{\theta}_i) + \bar{m}_i(\theta_i^k) + \bar{\epsilon} > u_i^*(\bar{\theta}_i)$, which contradicts *bribe-proofness.* □

Finally, note that the assumption of a compact type-space was only needed temporarily.

COROLLARY 1 *Suppose that Y is compact and φ is* bribe-proof*. Then u_j^* is continuous in Θ_i.*

Proof: Suppose not. Then u_j^* violates continuity on some compact subdomain, $\bar{\Theta}_i \subset \Theta_i$. The restriction of φ to the subdomain $\bar{\Theta}_i \times \Theta_{-i}$ defines a discontinuous *bribe-proof* solution (on a new domain) violating the conditions of Theorem 1. □

We end this section by noting that the assumption of a compact Y can not simply be dropped.

EXAMPLE 2 Let $Y = [0, \infty)$, $\Theta_1 = [0, 1]$, and v_1 satisfy

$$v_1(y, \theta_1) = \begin{cases} -y & \text{if } \theta_1 = 0, \\ \max\{-y, 2 - |\frac{1}{\theta_1} - y|\} & \text{if } \theta_1 > 0, \end{cases}$$

Let $\Theta_2 = \{\theta_2\}$, and $v_2(y; \theta_2) \equiv 0$. Let $\hat{\varphi}$ satisfy

$$\hat{\varphi}(\theta_1, \theta_2) = (\bar{y}(\theta_1), \bar{m}_1(\theta_1), \bar{m}_2(\theta_1)) = \begin{cases} (0, 0, 0) & \text{if } \theta_1 = 0, \\ (\frac{1}{\theta_1}, -\frac{1}{2}, \frac{1}{2}) & \text{if } \theta_1 > 0, \end{cases}$$

One may check that $\hat{\varphi}$ is *bribe-proof*, but that neither u_1^* nor u_2^* is continuous at $\theta_1 = 0$. A similar example can be constructed in which Y is bounded but open.

3.2 Finite Sets of Public Outcomes

Consider again the *bribe-proof* solutions in Example 1. Suppose that instead of being able to choose any point in some interval $[a, b]$, we may only choose integer points in the interval, e.g. $Y = \{1, 2, 3\}$. Median-voter types of solutions are well-defined in this setting also, subject to some tie-breaking procedure when the "median voter" is indifferent between two elements of Y. such solutions are, however, no longer *bribe-proof*! Informally, when the median voter is indifferent (or almost indifferent) between two elements of Y, he could be bribed to misreport his preferences. The reader is left to check this.

A trivial example of a *bribe-proof* solution for this environment is a constant solution. The discouraging news is that for connected domains in general, if Y is finite, then this is essentially the *only* type of *bribe-proof* solution. The following example illustrates why we say it is *essentially* so.

EXAMPLE 3 (TWO OUTCOMES) Consider two agents in neighboring offices, each having his own air conditioner which can be turned on or off. An agent's type represents the value of having the air conditioner in his own office turned on. Let $N = \{1, 2\}$, $Y = \{\text{On}, \text{Off}\}^2$, and $\Theta_1 = \Theta_2 = \mathbb{R}$. Furthermore, while agent 2 has a quiet air conditioner, agent 1 has a noisy one, which annoys agent 2 if it is turned on. So, the agents' valuation functions satisfy, for all $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$,

$$\begin{array}{ll} v_1((\text{On}, \cdot); \theta_1) = \theta_1 & v_2((\text{On}, \text{On}); \theta_2) = \theta_2 - 1 \\ v_1((\text{Off}, \cdot); \theta_1) = 0 & v_2((\text{On}, \text{Off}); \theta_2) = -1 \\ & v_2((\text{Off}, \text{On}); \theta_2) = \theta_2 \\ & v_2((\text{Off}, \text{Off}); \theta_2) = 0 \end{array}$$

The following "natural" solution, in which agent 2 has control over his air conditioner and agent 1 compensates agent 2 if he uses his, is *bribe-proof.*

$$\hat{\varphi}(\theta) = \begin{cases} ((\text{On}, \text{On}), -1, 1) & \text{if } \theta_1 \geq 1 \text{ and } \theta_2 \geq 0 \\ ((\text{On}, \text{Off}), -1, 1) & \text{if } \theta_1 \geq 1 \text{ and } \theta_2 < 0 \\ ((\text{Off}, \text{On}), 0, 0) & \text{if } \theta_1 < 1 \text{ and } \theta_2 \geq 0 \\ ((\text{Off}, \text{Off}), 0, 0) & \text{otherwise} \end{cases}$$

For the solution in Example 3, agent 2 plays the role of a dictator over the second dimension of the outcome space (the status of his own air conditioner). This is compatible with *bribe-proofness* because agent 1 has "trivial" preferences over that dimension—his preferences are unaffected by the status of that air conditioner. Similarly, agent 2 has trivial preferences over the first dimension of Y—regardless of his type, he always is one unit worse off when agent 1's air conditioner is on. Agent 1 plays the role of a dictator on that dimension, subject to the constraint that agent 2 be indifferent between any "choices" agent 1 has.

For connected domains and finite sets of outcomes, this is in fact a consequence of *bribe-proofness*—when the solution is responsive to a change in an agent's type, the change must not affect any other agent's payoff.

THEOREM 2 *Suppose that Y is finite, Θ is a connected domain, and φ is* bribe-proof. *Then u_j^* is constant in Θ_i.*

Proof: Since Y is finite, so is $\bar{O}_i$.[9] Therefore, O_i^* is finite, and the Maximum Theorem directly implies O_i^* is u.s.c. (without the assumption that Θ_i is compact, as needed above). Hence for all $\theta_i \in \Theta_i$ there exists $\delta > 0$ such that $|\theta_i' - \theta_i| < \delta$ implies $O_i^*(\theta_i') \subseteq O_i^*(\theta_i)$. Therefore for any such θ_i', we have $v_i(\bar{y}(\theta_i', \theta_{-i}); \theta_i) + \bar{m}_i(\theta_i', \theta_{-i}) = u_i^*(\theta_i)$. Therefore *bribe-proofness* implies $u_j^*(\theta_i') \leq u_j^*(\theta_i)$ (otherwise j would bribe i).

We have shown that each $\theta_i \in \Theta_i$ is a local maximizer of u_j^*. Therefore, since Θ_i is path-connected and u_j^* is continuous (Corollary 1, u_j^* is constant (see Lemma 2 in the Appendix). □

Theorem 2 implies that under a *bribe-proof* solution, each agent is actually a dictator on the range of the solution. Formally, the solution must satisfy the following condition.

***All-Dictatorial*:** The solution $\hat{\varphi} = (\bar{y}, \bar{m})$ is *all-dictatorial* if for all $\theta, \theta' \in \Theta$ and all $i \in N$, we have $v_i(\bar{y}(\theta); \theta_i) + \bar{m}_i(\theta) \geq v_i(\bar{y}(\theta'); \theta_i) + \bar{m}_i(\theta')$.

The Corollary follows directly from Theorem 2.

COROLLARY 2 *Suppose that Y is finite and Θ is a connected domain. Then φ is* bribe-proof *if and only if φ is* all-dictatorial.

A relevant concept here is Hurwicz and Walker's (1990) notion of a "decomposable" domain. When each agent cares only about his own dimension of the outcome space (i.e., the domain is decomposable), it is a trivial matter to define *bribe-proof* solutions. They can even be efficient, but they need not

[9]Recall that by *strategy-proofness*, $(y, m_i), (y, m_i') \in \bar{O}_i$ implies $m_i = m_i'$.

be anything close to "constant", as in Example 3. When there is a "conflict of interest" between agents (and the domain is indecomposable), however, the condition (*all-dictatorial*) is much stronger. For example, in Section 3.4 we examine a class of domains for which this conflict of interests always exists, and derive an even stronger conclusion.

The next result follows from Corollary 2.

Corollary 3 *Suppose that* Y *is finite,* Θ *is a connected domain, and* φ *is* bribe-proof. *For all* $\theta, \theta' \in \Theta$*, if* $\varphi(\theta) = (y, m)$ *and* $\varphi(\theta') = (y, m')$*, then* $m = m'$.

Before concluding this section, note that even if a domain of interest is not connected, the results could be applied to each "connected component" of the domain, that is, each subdomain that itself forms a connected domain. For example, if a 2-agent domain consisted of four connected components, Theorem 2 implies that under a *bribe-proof* solution, there are four pairs of payoffs, each associated with one of the components.

3.3 Smooth Preferences

Consider once again the solutions described in Example 1. Note that each single-peaked valuation function has a slope of 1 to the left of its peak and a slope of -1 to the right. As we did originally, let $Y = \mathbb{R}$, but now suppose that the domain was such that "steeper" and "flatter" single-peaked valuation functions were also admissible. Again, the median-voter types of solutions are well-defined in such a setting. They are not, however, *bribe-proof.* If the median voter has a relatively flat valuation function, he cares less about the location of y, relative to his transfer, than do agents with steeper valuation functions, hence he can be bribed. Similarly, if valuation functions were smooth, the same problem would arise: The median voter would have a "locally flat" valuation function at his peak, and could be bribed to make at least a small misrepresentation.

In fact, Theorem 2 and Corollary 2 generalize to the case in which Y is infinite, as long as the domain is "smoothly" connected, in the sense of Holmström (1979). That is, there should exist a smooth, one-dimensional parameterization of some path between any two types:

Smoothly Connected: The domain Θ is smoothly connected if for all $i \in N$ and all $\theta_i, \theta_i' \in \Theta_i$, there exists $w: Y \times [0,1] \to \mathbb{R}$ such that

i. For all $x \in [0,1]$, there exists $\theta_i^x \in \Theta_i$ such that $w(\,\cdot\,, x) = v_i(\,\cdot\,; \theta_i^x)$

ii. $w(\,\cdot\,, 0) = v_i(\,\cdot\,; \theta_i)$

iii. $w(\,\cdot\,, 1) = v_i(\,\cdot\,; \theta_i')$

iv. For all $y \in Y$, $w(y, \cdot\,)$ is differentiable on $[0,1]$

v. There exists $z \in \mathbb{R}$ such that for all $y \in Y$ and all $x \in [0,1]$,

$$\left| \frac{\partial w(y,x)}{\partial x} \right| \leq z$$

One may check that the domain of Example 1 can not be parameterized in this way, so it is not smoothly connected. On the other hand, if v_i is differentiable in Θ_i for all $i \in N$, then any convex Θ is smoothly connected.

Theorem 3 *Suppose that Θ is a smoothly connected domain and φ is* bribe-proof*. Then u_j^* is constant in Θ_i.*

Proof: As in the definition of smoothly connected domains, let w be defined with respect to Θ_i, and for all $x \in [0,1]$, let θ_i^x be defined as in (i).

Define $f: [0,1]^2 \to \mathbb{R}$ so that for all $x, x' \in [0,1]$,

$$f(x, x') \equiv v_i(\bar{y}(\theta_i^x, \theta_{-i}), \theta_i^{x'}) + \bar{m}_i(\theta_i^x, \theta_{-i})$$

Strategy-proofness implies that for all $x' \in [0,1]$,

$$x' \in \arg\max_{x \in [0,1]} f(x, x') \tag{1}$$

Bribe-proofness implies that for all $x' \in [0,1]$,

$$x' \in \arg\max_{x \in [0,1]} f(x, x') + u_j^*(\theta_i^x) \tag{2}$$

Since Θ is smoothly connected, the Lemma in the Appendix of Holmström (1979) states that (1) and (2) imply that u_j^* is constant. □

As in the previous section, this result can be used to derive the following.

Corollary 4 *Suppose that* Θ *is a smoothly connected domain. Then* φ *is* bribe-proof *if and only if* φ *is* all-dictatorial.

Corollary 5 *Suppose that* Θ *is a smoothly connected domain and* φ *is* bribe-proof. *For all* $\theta, \theta' \in \Theta$, *if* $\varphi(\theta) = (y, m)$ *and* $\varphi(\theta') = (y, m')$, *then* $m = m'$.

3.4 Rich Domains

One might observe that in the above examples in which non-constant *bribe-proof* solutions exist, the domains are, loosely speaking, "narrow". For example, they don't contain many perturbations of the functions they contain. It turns out that if a domain is rich enough, then in fact only constant functions are *bribe-proof*.

We will use the following definition of richness, which requires that if a valuation function is admissible, then for any outcome $y \in Y$, there exists another admissible valuation function for which the value of y, relative to any other outcome, is strictly greater than for the original function.

Monotonically Closed: The domain Θ is monotonically closed if for all $i \in N$, all $\theta_i \in \Theta_i$, and all $y \in Y$, there exists $\theta_i' \in \Theta_i$ such that for all $y' \in Y \setminus \{y\}$, $v_i(y; \theta_i') - v_i(y'; \theta_i') > v_i(y; \theta_i) - v_i(y'; \theta_i)$.

Theorem 4 *Suppose that* Θ *is connected and monotonically closed. Further, suppose that either* Y *is finite or* Θ *is smoothly connected. If* φ *is* bribe-proof, *then it is a constant function.*

Proof: We will prove the theorem for the case in which Y is finite by using the results of Section 3.2. The proof for the case in which Θ is smoothly connected is the same but uses the results of Section 3.3.

Suppose by contradiction that there exist distinct (y, m) and (y', m') in the range of φ. By Corollary 3, $y \neq y'$, so without loss of generality we have $(y, m_i), (y', m_i') \in \bar{O}_i$ with $y \neq y'$. In this proof we will change agent j's type (from θ_j). To simplify notation, let $\varphi = (\bar{y}, \bar{m})$ depend only on the types of agents i and j.

Let $\theta_i \in \Theta_i$ satisfy $\varphi(\theta_i, \theta_j) = (y, m)$. Since agent i receives (y', m_i') for *some* reported type, and since Θ is monotonically closed, there exists $\theta_i' \in \Theta_i$

such that

$$\{(y', m_i')\} = \underset{(\hat{y},\hat{m}_i)\in\bar{O}_i}{\arg\max}\, v_i(\hat{y};\theta_i) + \hat{m}_i \tag{3}$$

Strategy-proofness implies $\varphi_i(\theta_i', \theta_j) = (y', m_i')$. Corollary 3 therefore implies

$$\varphi(\theta_i', \theta_j) = (y', m') \tag{4}$$

Corollary 2 implies that for all $\hat{\theta}_i, \hat{\theta}_j$, $v_j(y;\theta_j) + m_j \geq v_j(\bar{y}(\hat{\theta}_i, \hat{\theta}_j);\theta_j) + \bar{m}_j(\hat{\theta}_i, \hat{\theta}_j)$. So, since Θ is monotonically closed, there exists $\theta_j' \in \Theta_j$ such that for all $\hat{\theta}_i, \hat{\theta}_j$,

$$\varphi_j(\hat{\theta}_i, \hat{\theta}_j) \neq (y, m_j) \Longrightarrow v_j(y;\theta_j') + m_j > v_j(\bar{y}(\hat{\theta}_i, \hat{\theta}_j);\theta_j) + \bar{m}_j(\hat{\theta}_i, \hat{\theta}_j) \tag{5}$$

Strategy-proofness then implies $\varphi_j(\theta_i, \theta_j') = (y, m_j)$, so by Corollary 3,

$$\varphi(\theta_i, \theta_j') = (y, m) \tag{6}$$

Theorem 2 implies

$$v_j(\bar{y}(\theta_i', \theta_j');\theta_j') + \bar{m}_j(\theta_i', \theta_j') = v_j(\bar{y}(\theta_i, \theta_j');\theta_j') + \bar{m}_j(\theta_i, \theta_j')$$

With eqns. (5) and (6), this implies $\varphi_j(\theta_i', \theta_j') = (y, m_j)$, so by Corollary 3,

$$\varphi(\theta_i', \theta_j') = (y, m) \tag{7}$$

Similarly, Theorem 2 implies

$$v_i(\bar{y}(\theta_i', \theta_j');\theta_i') + \bar{m}_i(\theta_i', \theta_j') = v_i(\bar{y}(\theta_i', \theta_j);\theta_i') + \bar{m}_i(\theta_i', \theta_j)$$

With eqns. (4) and (7), this implies $v_i(y;\theta_i') + m_i = v_i(y';\theta_i') + m_i'$, contradicting eqn. (3). □

4 Conclusion

We have presented a generalized model in which agents have quasi-linear preferences over outcomes and transfers, and shown that in many situations (domains), it is essentially impossible to design a non-trivial solution immune to manipulation by pairs of agents when they can make transfers. The specification of the model is such that the simplest interpretation of an outcome is as a public decision, such as a level of public goods. Through the correct specification of a domain, however, there are applications of this model to many other environments. Examples include auctions, more general allocation problems with indivisible goods,[10] matching problems with money, queuing problems, exchange economies with or without production, and voting models.[11]

Our concept of manipulation is defined in such a way as to rule out situations in which two agents *with given types* could manipulate a solution. This condition is most readily applied to situations in which agents have complete information about each others' types. It should be noted, however, that the results apply even to cases in which agents have partial information.[12] For example, it may be the case that agents only know the types of their "neighbors." In such a case, our results apply to the payoff of an agent (only) with respect to changes in his neighbors reported type. As another example, it may be the case that an agent can identify a second agent's type only when that type is in some given subset of types. Again in such a case, our results apply locally, regarding the agent's payoff with respect to changes in the second agent's type *within that subset* of types. We leave the formalization of such results to the reader.

Of additional interest is a result by Crémer (1996) regarding the manipulability of Clarke–Groves mechanisms by groups of agents for the case of exactly two outcomes ($|Y| = 2$). The remainder of the section is dedicated to an informal discussion of Crémer's result for groups of two agents, and the way our two sets of results together establish a boundary between possibility and impossibility at the point where agents gain information about each others' types.

Crémer's setup is as follows. Imagine that a Clarke–Groves mechanism is being used, and that all of the agents except, say, i and j have already reported their types to the mechanism (so interpret them as fixed). Further,

[10] See Schummer (1999) for a version of Theorem 2 in such an environment.

[11] For example, to model a simple 2-good exchange economy in which preferences are quasi-linear in the first good, the set of "public outcomes" is a simplex representing the division of the second good among the agents.

[12] I thank a referee for drawing attention to this point.

imagine that agents i and j do not know each other's types, but anticipate the possibility of gains by *jointly* misrepresenting their types *and* making an internal transfer. Since they do not know each others' types, they coordinate their potential misrepresentation by devising a "sub-mechanism", to which they report their types, and which determines for them (i) a (mis-)report of their types to be made to the original Clarke–Groves mechanism, and (ii) a transfer to be made between the two.

The question is whether a pair of agents could devise such a sub-mechanism that is itself *strategy-proof*. The answer is *sometimes*: Crémer (1996) provides some Clarke–Groves mechanisms that are immune to such manipulation by pairs of agents.[13]

One may think that Theorem 2 contradicts Crémer's result with the following reasoning: If a Clarke–Groves mechanism is not *bribe-proof*, as shown by Theorem 2, then why can we not design a sub-mechanism for some pair of agents, as above, to implement this bribe, violating the result of Crémer? The answer is that such a sub-mechanism would be manipulable by one of the two agents—one of the two agents cheating the system will be cheated by the other agent.

For a precise example, consider a Clarke–Groves mechanism, which is not *bribe-proof*: For some types $\theta \in \Theta$, agent j can successfully bribe agent i with b units of money to mis-report his type as θ'_i. In order to take advantage of this, one may propose the following Crémer-style sub-mechanism for the two agents: Given that the other agents have reported types θ_{-ij}, whenever i and j report (θ_i, θ_j) to the sub-mechanism, the sub-mechanism recommends the mis-report (θ'_i, θ_j) plus a transfer of b to be made from j to i. In all other cases, the sub-mechanism recommends no misrepresentation and no transfer.

Is this sub-mechanism *strategy-proof*? It is simple to check that in most cases, this sub-mechanism can not be manipulated. However, when their types are (θ'_i, θ_j), agent i can mis-report to the sub-mechanism that he is of type θ_i; the sub-mechanism then recommends the "mis-report" (θ'_i, θ_j) to the Clarke–Groves mechanism (which is what i would have reported anyway) *plus* a transfer of b to agent i, resulting in a gain to agent i. Hence this is not a *strategy-proof* sub-mechanism.

[13] In turn, he shows that all Clarke–Groves mechanisms are manipulable to this sort of manipulation by *triples* of agents. On the other hand, he shows that each such manipulation by a triple is, itself, re-manipulable by two of those agents!

More generally, Crémer's result tells us that for a particular class of Clarke–Groves mechanisms, there is no *strategy-proof* sub-mechanism that allows a pair of agents to take advantage of any such bribing situation.

The essential difference between *bribe-proofness* and Crémer's manipulation condition is in the need for agents to know each other's types. Under the stronger condition of *bribe-proofness*, a manipulation is considered possible if there is any situation in which a pair of agents could gain through the bribing procedure—this applies most readily to situations in which agents have information (see above) about each others' types. But for a pair of agents to be able to gain with a sub-mechanism, they must devise a plan of manipulation that covers all possible realizations of their types, and it must be immune to further manipulation by any of the two individuals.

Since the two concepts are similar except for their respective implicit assumptions regarding the information agents have about each others' preferences, the two sets of results could be seen as a dividing line between the possibility and the impossibility of having solutions that are non-manipulable by coalitions (or at least pairs) of agents. Possibility obtains even among the class of Clarke–Groves mechanisms as soon as potentially misrepresenting agents lose the information of each other's types in a simple public goods environment. With more information, however, manipulation is possible under almost any solution, in many different kinds of environments.

Appendix

Lemma 2 *Let X be a path-connected set. If $f\colon X \to \mathbb{R}$ is continuous and if for all $x \in X$, x is a local maximizer of f, then f is constant.*

Proof: Suppose f is continuous and not constant. Then there exist $x, y \in X$ such that $f(x) < f(y)$. Let $g\colon [0,1] \to X$ be continuous, and satisfy $g(0) = x$ and $g(1) = y$. Let $L = \{\delta \in [0,1] : 0 \leq \delta' \leq \delta \Longrightarrow f(g(\delta')) \leq f(x)\}$. (Note that L is a non-empty, connected set.) Let $\bar{\delta} \equiv \sup L$.

Since f is continuous, $\bar{\delta} \in L$, so $g(\bar{\delta})$ is not a local maximizer of f. □

The continuity of f can be replaced with every x also being a local minimizer: If f is not continuous, then $g(\bar{\delta})$ in the proof of the Lemma is either not a local maximizer or not a local minimizer. That is, if f is not constant, then either there exists a non-local-maximizer or there exists a non-local-minimizer, but not necessarily both. For example, consider the function $f(x) = 0$ for $x \neq 1$ and $f(1) = 1$; every x is a local maximizer.

Proof of Corollary 2

COROLLARY 2 *Suppose that Y is finite and Θ is a connected domain. Then φ is* bribe-proof *if and only if φ is* all-dictatorial.

Proof: Suppose by contradiction that for some $\theta, \theta' \in \Theta$ and $k \in N$,

$$v_k(\bar{y}(\theta); \theta_k) + \bar{m}_k(\theta) < v_k(\bar{y}(\theta'); \theta_k) + \bar{m}_k(\theta')$$

By repeated application of Theorem 2,

$$v_k(\bar{y}(\theta_k, \theta'_{-k}); \theta_k) + \bar{m}_k(\theta, \theta'_{-k}) < v_k(\bar{y}(\theta'); \theta_k) + \bar{m}_k(\theta')$$

contradicting *strategy-proofness.* □

Coalitional Strategy-proofness

It is simple to observe that any *all-dictatorial* solution is also *coalitionally strategy-proof.* Therefore on any domain of the types discussed in Sections 3.2–3.4, *bribe-proofness* implies *coalitional strategy-proofness.* The following trivial example shows, however, that this logical relation does not hold in general.

TABLE I
Definition of valuation functions.

	a	b	c	d
$v_1(\,\cdot\,; 0)$	0	2	−10	1
$v_1(\,\cdot\,; 1)$	1	−10	2	0
$v_2(\,\cdot\,; 0)$	0	−10	2	1
$v_2(\,\cdot\,; 1)$	1	2	−10	0

Let $N = \{1, 2\}$, $Y = \{a, b, c, d\}$, and $\Theta_1 = \Theta_2 = \{0, 1\}$. Let the valuation functions be defined as in Table I.

One *bribe-proof* solution that is not *coalitionally strategy-proof* is $\varphi = (\bar{y}, \bar{m})$, defined by $\bar{m}_1(\cdot) \equiv 0 \equiv \bar{m}_2(\cdot)$, and

$$\begin{aligned} \bar{y}(0,0) &= a \\ \bar{y}(0,1) &= b \\ \bar{y}(1,0) &= c \\ \bar{y}(1,1) &= d \end{aligned}$$

Also note that this rule is not efficient; hence *bribe-proofness* does not even imply *efficiency*.

References

Barberà, S. and M. Jackson (1995): "Strategy-proof Exchange," *Econometrica*, 63, 51–87.

Berge, C. (1963): *Topological Spaces*. New York: Macmillan.

Chichilnisky, G. and G. Heal (1997): "The Geometry of Implementation: A Necessary and Sufficient Condition for Straightforward Games," *Social Choice and Welfare*, 14, 259–94.

Clarke, E. (1971): "Multipart Pricing of Public Goods," *Public Choice*, 8, 19–33.

Crémer, J. (1996): "Manipulations by Coalitions Under Asymmetric Information: The Case of Groves Mechanisms," *Games and Economic Behavior* **13**, 39–73.

Green, J. and J.-J. Laffont (1977): "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, 45, 427–38.

——— (1979): "On Coalition Incentive Compatibility," *Review of Economic Studies*, 46, 243–54.

Groves, T. (1973): "Incentives in Teams," *Econometrica*, 41, 617–31.

Holmström, B. (1979): "Groves' Scheme on Restricted Domains," *Econometrica*, 47, 1137–44.

Hurwicz, L. (1972): "On Informationally Decentralized Systems," in *Decision and Organization*, ed. by C. B. McGuire and R. Radner. Amsterdam: North Holland, 297–336.

Hurwicz, L. and M. Walker (1990): "On the Generic Non-optimality of Dominant Strategy Allocation Mechanisms," *Econometrica*, 58, 683–704.

Moulin, H. (1980): "On Strategy-proofness and Single-peakedness," *Public Choice*, 35, 437–55.

Schummer, J. (1999): "Eliciting Preferences to Assign Positions and Compensation," forthcoming in *Games and Economic Behavior*.

Serizawa, S. (1998): "Pairwise Strategy-Proofness," mimeo, Shiga University.

Sundaram, R. (1996): *A First Course in Optimization*, New York: Cambridge University Press.

Tirole, J. (1992): "Collusion and the Theory of Organizations," in *Advances in Economic Theory, Sixth World Congress*, ed. by J.-J. Laffont. Cambridge, U.K.: Cambridge University Press.