Almost-dominant Strategy Implementation

James Schummer*

MEDS Department, Kellogg Graduate School of Management

Northwestern University, Evanston, IL 60208–2009

Email: schummer@kellogg.northwestern.edu

First version: August, 1999

Revised: December 10, 2002

Abstract

Though some environments yield reasonable allocation rules that are implementable in dominant strategies (i.e., strategy-proof), a significant number yield impossibility results. On the other hand, while there are general possibility results for implementation in Nash or Bayesian equilibrium, these equilibrium concepts make strong assumptions about the players' knowledge. Since such assumptions may not be practical in various design scenarios, we formulate a solution concept built on one premise: Players who do not have much to gain by manipulating will not bother to do so.

For an exchange economy model and a voting/lotteries model, we search for efficient rules that never provide players with large gains from manipulation. Though the rules we describe are inequitable, they are significantly more flexible than those that satisfy the stronger condition of strategy-proofness, even when the allowable gains from manipulation are made arbitrarily small. This demonstrates a type of non-robustness in previous impossibility results.

Keywords: strategy-proof, almost dominant strategy, ϵ-dominant.

JEL Classification Numbers: C70, D70.

*I thank Ehud Kalai for very helpful comments.

1 Introduction

In the field of mechanism design, one of the most desirable incentives properties for choice rules is that of dominant strategy implementability. For a planner attempting to implement a rule with this property, certain issues—such as what information he has about the agents, what information agents have about each other, and what information is revealed during intermediate stages of the execution of the mechanism—are basically irrelevant. These issues are also irrelevant for a participating agent calculating an appropriate (best) action to take. In fact, even the assumption that his fellow players are rational need not be made by the player concerned with his own best interests. Furthermore, calculating a player's best action cannot be more complex than determining his own preferences over outcomes.

Given the extreme desirability of this incentives property, it is important to determine which rules satisfy it in various situations. Indeed this question has been—and continues to be—answered for an increasingly diverse class of situations. Interestingly, though, the nature of the result depends strongly on the situation being described.

For example, the seminal works of Gibbard (1973) and Satterthwaite (1975) provided an early negative result for the situation of voting: no (non-dictatorial) voting rule satisfies even the (slightly weaker) condition of *strategy-proofness*, requiring truth-telling to be *a* weakly dominant strategy in the direct revelation mechanism. For the situation of choosing public alternatives and taxation levels, the Vickrey–Clarke–Groves[1] mechanisms have been shown (e.g., by Green and Laffont (1977), Holmström (1979)) to be the only *strategy-proof* ones that choose efficient public alternatives. Since these mechanisms are typically not budget balancing, this has been seen as a negative result.[2]

In contrast, in certain "simpler" situations, positive results have prevailed. Moulin (1980) describes the class of *strategy-proof*, onto voting rules for the situation in which agents have "single-peaked" preferences over a 1-

[1]See Clarke (1971) and Groves (1973).

[2]However, this class of mechanisms has a much better reputation in private goods environments.

dimensional ordering of public alternatives, generalizing the classic median-voter rule.[3] For two classes of 2-sided matching problems (known as marriage markets and college admissions problems), Alcaldé and Barberà (1994) provide a domain of preferences for which certain stable matching rules are *strategy-proof.* In such problems, stability is arguably the most important property for a rule to possess. For 1-sided matching problems—in particular, Shapley and Scarf's (1974) "housing market"—Roth (1982) demonstrates the *strategy-proofness* of the allocation rule that has been central to the analysis of this domain: the Top Trading Cycles algorithm. For situations in which agents have single-peaked preferences over consumption of a single, divisible private good, Sprumont (1991) shows the *strategy-proofness* of the Uniform Rule, which has subsequently been characterized in terms of many other desirable properties; see Ching (1992,1994), Schummer and Thomson (1997), and Thomson (1994a,b,1995).

As the literature on *strategy-proofness* grows, our picture of the dividing line between possibility and impossibility becomes clearer.[4] This leaves us with the need to address those situations in which no reasonable rules are implementable in dominant strategies. There are various ways to do this.

One approach is to require a weaker form of implementation. This is the approach taken in the large literature on *Nash implementation* (and its refinements), in which mechanisms have the property that their equilibrium outcomes are ones that would have been chosen by some given choice rule. For example, see Moore (1996). The results here tend to be more positive than those in the *strategy-proofness* literature.

However, these results come with a price: Strong assumptions are made concerning the structure of information that agents possess about each other, and/or that the planner possesses about the players. For example, it is implicitly assumed in this literature that players have either common knowledge or common prior beliefs about each other's preferences). For many mechanism design environments, such an assumption is not realistic.

[3]This result is strengthened by Ching (1997) and generalized by Schummer and Vohra (2001).

[4]For a more detailed survey of the *strategy-proofness* literature, including more positive results, see Barberà (2001) and Thomson (1998).

A second, related approach to addressing the impossibilities of *strategy-proofness* applies to situations in which the planner is satisfied with approximations; he may find it sufficient to implement a rule that is "close" to some other desirable choice rule. One application of this approach can be seen in the literature on *virtual implementation* (Abreu and Matsushima (1992), Duggan (1997)), in which the goal is to find an implementable mechanism whose equilibrium outcomes approximate the desired outcomes. The solution concepts in this literature, however, make the same type of informational assumptions listed above.

Another application of the "approximation approach" is to measure, in some way, the manipulability of a mechanism. In fact, there are various ways of performing *this* analysis. One way, which is applicable to large economies, is to look at the asymptotic behavior of incentives conditions. Roberts and Postlewaite (1976) observe that as the number of agents becomes large, the Walrasian allocation rule is asymptotically *strategy-proof* (see also Córdoba and Hammond (1998) and Ehlers et al. (1999)).

A similar analysis in an auction setting is performed by Rustichini, Satterthwaite, and Williams (1994), and Satterthwaite (2001), who not only show an analogous asymptotic result for double auctions, but also argue that convergence happens quickly as the number of agents increases. Conversely, Swinkels (2001) shows that the Nash equilibrium outcomes of various popular auction mechanisms are asymptotically efficient.

A related asymptotic result is given by Kalai (2001), who shows that in many anonymous games of incomplete information, Bayesian equilibria yield, with high probability, ex-post ϵ-equilibrium outcomes.

Alternatively, Beviá and Corchón (1995) consider measuring the frequency of manipulation opportunities. They show that in a public goods setting, any efficient and individually rational mechanism must be manipulable on a dense set of preference profiles. Kelly (1993), Saari (1995), and Smith (1999) suggest ways of counting manipulable situations in a discrete voting environment.

Finally, Harrison and McDaniels (2001) argue (with experimental evidence) that since a certain Condorcet-consistent voting rule is computationally difficult to manipulate, it is not likely to be manipulated in practice.

Our Approach

The approach taken in this paper can be seen as a different type of contribution to the approximation approach, involving an approximation to the notion of a dominant strategy. The motivation behind our notion of approximation lies with a simple assumption about the strategic behavior of agents. Specifically, we approach the problem with the single premise that if a player does not have much to gain by manipulating an allocation rule, then he will not bother to manipulate it. Under this modelling assumption, we search for rules in which gains from manipulation are limited by an upper bound.

This assumption can be interpreted or applied in various ways. For example, it applies when agents must incur a cost in order to gather information about each other. If such costly information is necessary for a player to compute a profitable way to manipulate the choice rule, it would never be worth the expense to gather it if the potential gains were bounded above by this cost. Another application of this idea is to situations in which the act of computation itself is costly to a player.

A third example of an application of our assumption is to situations in which agents value morality (or honesty) in some real, fixed terms. In such settings, small gains from cheating do not outweigh the losses (or "guilt") incurred.

An important observation here is that we make no assumption on the structure of information that agents possess. Some of the work cited previously considers a rule to be almost non-manipulable in a Bayesian setting even if there is a small probability of a very large gain. Such a definition implicitly assumes that players not only have beliefs consistent with those assumed by the planner, but that the players *cannot* have *more* information than that.

The hypothesis of our premise is that "a player does not have *much* to gain by manipulating." The critical detail of our work is to precisely define *much*. One approach is to use a utility-based approach to preferences. Using this approach, a player would be assumed not to manipulate a rule unless his utility gain would exceed a predetermined amount. This approach, however, would depend heavily on the interpretation (and/or the parameterization) of utility functions.

To avoid this difficulty, we define our condition in terms of commodities. For example, we first examine 2-agent exchange economies. In such a model, our behavioral assumption is that the only situations in which an agent will manipulate a choice rule are those situations when his gains are better than receiving prespecified, additional amounts (say, ϵ_j) of any good j. If no such situation exists, we say that the choice rule is "almost" *strategy-proof*.

Second, we examine a voting model in which outcomes are lotteries over candidates. Here, our behavioral assumption is that an agent will manipulate a choice rule only if he can obtain a lottery which (from his perspective) is better than any other lottery within an ϵ-neighborhood of the original lottery. The "commodity space" in this model is the simplex of lotteries.

Overview of Results

Our first model is of a 2-agent exchange economy with two goods. We restrict attention to the domain of linear (additively separable) preferences. We begin with this simple class of preferences for two reasons: the analysis is more tractable, and it is straightforward to quantify certain bounds imposed by the truth-telling condition (i.e., as ϵ is increased), as we discuss below. In any case, the positive aspect of our results can be extended to other domains, as discussed in Section 7.

One of the earlier works on mechanism design is a paper by Hurwicz (1972), concerning 2-agent exchange economies with a more general domain of preferences.[5] He shows that it is impossible to construct a *strategy-proof*, efficient rule that provides allocations which both agents prefer to their original endowment. Zhou (1991) improves upon this result by showing that if a rule is *strategy-proof* and efficient, then it is dictatorial: it must always give all of the goods to a prespecified agent. Finally, Schummer (1997) strengthens these results by showing them to hold even on "small" domains of preferences, including the linear preferences we use here.

[5] See also, Barberà and Jackson (1995).

Our results show that when *strategy-proofness* is weakened by any amount as described above, a larger class of rules becomes admissible. The class of newly admissible rules depends on the way in which we define our condition.

In Section 4, we consider all measures of gains to be made only with respect to a single good. Under this relatively stronger version of our condition, an efficient rule always allocates almost all of that good to a prespecified agent. On the other hand, the rule's allocation of the second good may have any degree of variability. This flexibility of rules is in strong contrast to the negative results cited above. Furthermore, we show in Section 4.3 that as our almost-dominance condition approaches *strategy-proofness* (i.e., as ϵ converges to zero), the ranges of admissible rules do *not* converge to the range of the only *strategy-proof*, efficient rules. That is, even if *strategy-proofness* is relaxed an arbitrarily small amount, there is a (discontinuous) increase in the flexibility of admissible rules.

In Section 5, we allow measures of gains to be made with respect to any good. Under this weaker version of almost-dominance, even more rules are admissible, and another discontinuity occurs as discussed in Section 5.1.

In Section 6, we parameterize the preferences of the agents, and quantify the effects of relaxing *strategy-proofness* to our almost-dominance condition. Under *strategy-proofness*, one agent must always consume nothing. Under the rules we discuss, one of the agents consumes relatively less than the other agent, but occasionally consumes a bundle of goods that he considers to be almost as good as the entire endowment. After choosing a particular class of utility functions to represent the preferences, we are better able to analyze this difference.

Our second model is that of choosing lotteries over public outcomes when agents have von Neumann–Morgenstern preferences. In this setting, Gibbard (1977) shows that if a *strategy-proof* and (ex ante) efficient rule depends only on the agents' preferences over degenerate lotteries, then it must always choose a prespecified agent's favorite (non-random) outcome.[6] Hylland (1980) strengthens Gibbard's result by dropping the latter informational requirement.

[6]Barberà (1977) obtains a related result when choosing sets of public outcomes.

In Section 8 we show, as in the previous model, that relaxing *strategy-proofness* to an almost-dominance condition leads to a discontinuous increase in the range of admissible rules. We show this by exploiting the geometric similarity of this environment to the previous one. In particular, we are able to reinterpret an important rule characterized in the exchange economy model to the lotteries model. The rule is not equitable—it always chooses one of the agent's "almost-favorite" outcomes—but it shows that the previous impossibility results are not entirely robust.

To summarize, the paper is organized as follows. In Section 2 we formalize the exchange economy model, while we provide our main definition in Section 3. In Sections 4 and 5, we provide our results for exchange economies. Using these results, we quantify the consequences of relaxing *strategy-proofness* in this model in Section 6. In Section 7, we provide a brief discussion on extending the analysis to more general models of exchange. In Section 8, the previous analysis is extended to the voting model with lotteries. Section 9 concludes.

2 Exchange Economy Model

The set of two agents is $N = \{1, 2\}$. There is a positive endowment of two infinitely divisible goods $\Omega = (\Omega^1, \Omega^2) \in \mathbb{R}^2_{++}$. Each agent $i \in N$ is to consume a bundle $x_i \in \mathbb{R}^2_+$. An *allocation* is a pair of bundles $x = (x_1, x_2) = ((x_1^1, x_1^2), (x_2^1, x_2^2)) \in \mathbb{R}^4_+$ such that $x_1 + x_2 = \Omega$; the set of allocations is denoted A. Subscripts refer to agents, superscripts refer to goods, and the vector inequalities are $>$, $\geq$, and $\geqq$.

Each agent has a strictly monotonic, linear preference relation, R_i, over his consumption space $\mathbb{R}^2_+$. Precisely, such preference relations are the ones representable by a utility function of the form $u(x_i) = \lambda x_i^1 + (1 - \lambda) x_i^2$, $\lambda \in (0, 1)$. Denote the set of such preference relations as $\mathcal{R}$. The strict (antisymmetric) and indifference (symmetric) preference relations associated with R_i are denoted P_i and I_i.

An *allocation rule* is a function, $\varphi \colon \mathcal{R}^2 \to A$, mapping the set of preference profiles into the set of allocations. To simplify notation, when $\varphi(R) = x$, we denote $\varphi_i(R) = x_i$ for any agent $i \in N$. Furthermore, we write $-i$ to refer to the agent not equal to i. For example, if $i = 1$, then $x_{-i} = x_2$, and (R'_i, R_{-i}) is the same as (R'_1, R_2).

We are interested in finding allocation rules that satisfy desirable properties not only in terms of incentives, but also in terms of efficiency. An allocation $x \in A$ is *efficient* with respect to a preference profile $R \in \mathcal{R}^2$ if there exists no $y \in A$ such that for some $i \in N$, $y_i \; P_i \; x_i$ and $y_{-i} \; R_{-i} \; x_{-i}$. We also call an allocation rule *efficient* if it assigns to every preference relation an allocation that is efficient with respect to that preference relation.

For any profile $R \in \mathcal{R}^2$, denote the set of efficient allocations for R as $E(R)$. On our domain of linear preferences, if both agents have the same preference relation $(R_1 = R_2)$, then the set of efficient allocations is the entire set: $E(R) = A$. If R is such that agent 1 values good 1 relatively more than agent 2 does, then the set of efficient allocations is $E(R) = E^{\lrcorner} \equiv \{x \in A : x_1^2 = 0 \text{ or } x_2^1 = 0\}$. In the opposite, remaining case, $E(R) = E^{\ulcorner} \equiv \{x \in A : x_1^1 = 0 \text{ or } x_2^2 = 0\}$.

3 A Definition of Nonmanipulability

Our goal is to find allocation rules that never afford agents the opportunity to gain much. The difficulty in formalizing this notion is to define the idea of *much* for all possible preference relations. Our approach will be to restrict attention to measures on the consumption space.[7] There are many ways to construct such measures of gains. In this section, we give one.[8]

A simple way to measure manipulability is to measure gains relative to either of the two goods; such a definition can be generalized to the case of more than two goods in an obvious way. To be precise, consider a situation in which an allocation rule φ prescribes, for $R \in \mathcal{R}^2$, an allocation $x = \varphi(R)$. If we postulate that agent i would not falsely report his preferences for small gains, then there exists some number $\epsilon_1 \geq 0$ such that if for some $R'_i \in \mathcal{R}$, we have $\varphi_i(R'_i, R_{-i}) \leqq \varphi_i(R) + (\epsilon_1, 0)$, then agent i would not manipulate the

[7] Alternatively, one could measure gains from manipulation in terms of some utility measure. However, we wish to avoid the implicit assumptions imposed by such modelling.

[8] There are definitions of manipulability similar to ours that yield results similar to the ones in this paper. We omit the exhaustive task of listing such definitions; it seems that there is no added insight from considering such similar definitions.

rule with that particular misrepresentation R'_i. That is, if agent i can gain only ϵ_1 (or fewer) units of good 1, then the gain is too small to be considered.

Similarly, for some $\epsilon_2 \geq 0$, we say that an agent does not manipulate φ if he simply gains ϵ_2 (or fewer) units of good 2.[9]

Finally, consider a situation in which a false report of preferences, R'_i, gives agent i the bundle $x_i = \varphi(R'_i, R_i)$, such that $\varphi_i(R) + (\epsilon_1, 0)\ R_i\ x_i$. Since the agent would not manipulate the rule in order to obtain the bundle $\varphi_i(R) + (\epsilon_1, 0)$, we conclude that he would not manipulate the rule in order to obtain the (worse) bundle x_i. A similar reasoning is to be applied with respect to good 2 and ϵ_2.

Our formal definition of this reasoning is as follows.

($\boldsymbol{\epsilon_1, \epsilon_2}$)-*strategy-proofness*: For any $\epsilon \in \mathbb{R}^2_+$, a rule is (ϵ_1, ϵ_2)-*strategy-proof* if for all $R \in \mathcal{R}^2$, all $i \in \{1, 2\}$, and all $R'_i \in \mathcal{R}$, we have either

(i) $\varphi_i(R) + (\epsilon_1, 0)\ R_i\ \varphi_i(R'_i, R_{-i})$, or

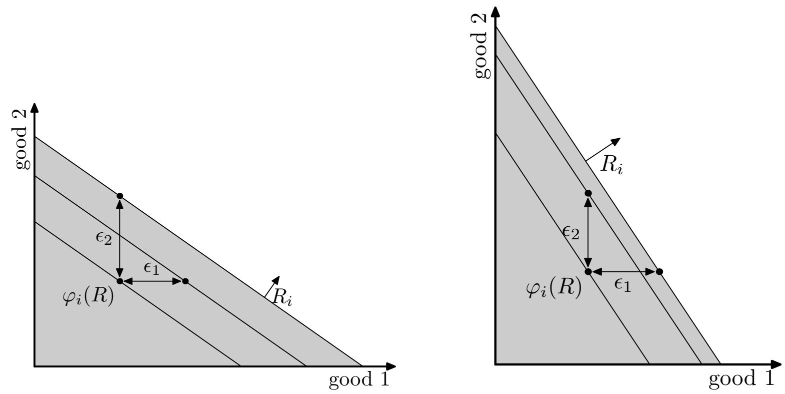
(ii) $\varphi_i(R) + (0, \epsilon_2)\ R_i\ \varphi_i(R'_i, R_{-i})$.

That is, by misreporting his preferences, an agent cannot procure a gain that he considers, simultaneously, to be (i) better than simply acquiring an additional ϵ_1 units of good 1 and (ii) better than simply acquiring an additional ϵ_2 units of good 2. See Figure 1; in Figure 1a, part [i] of the definition is redundant, while in the case of Figure 1b, part [ii] is. In the language of Barberà and Peleg (1990), the agent's *option set* should lie within the shaded area.

It should be clear that (ϵ_1, ϵ_2)-*strategy-proofness*is a stronger condition than $(\epsilon'_1, \epsilon'_2)$-*strategy-proofness* whenever $\epsilon \leq \epsilon'$, and that $(0, 0)$-*strategy-proofness* is equivalent to the standard definition of *strategy-proofness.*

We close this section by observing a previous result for the case $\epsilon = (0, 0)$. With a result related to that of Zhou (1991), Schummer (1997) shows that on this class of problems, the only efficient rules that are $(0, 0)$-*strategy-proof* are those that assign the entire endowment to a given agent.

[9] A more general definition makes the value of ϵ_j dependent on the identity of the agent in question, or, even more generally, his preference relation R_i. For simplicity, we do not go to this level of generality. See footnote 8.



(a) (b)

Figure 1: If φ is (ϵ_1, ϵ_2)-*strategy-proof*, then any false report by agent i results in a bundle somewhere within the shaded area.

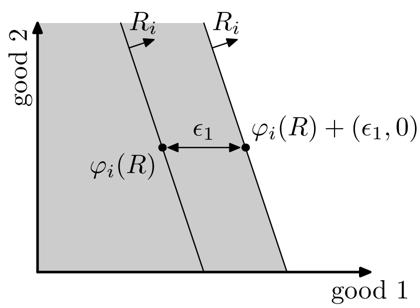


Figure 2: (ϵ_1, ϵ_2)-*strategy-proofness* when $\epsilon_2 = 0$.

Theorem 1 (Schummer (1997)) *Let* φ *be an efficient rule that is* $(0,0)$-strategy-proof*. There exists an agent* $i \in N$ *that always receives the entire endowment: for all* $R \in \mathcal{R}^2$*,* $\varphi_i(R) = (\Omega^1, \Omega^2)$ *(and* $\varphi_{-i}(R) = (0,0)$*).*

4 Results for $(\epsilon_1, 0)$-*strategy-proofness*

We first examine the implications of (ϵ_1, ϵ_2)-*strategy-proofness* when $\epsilon_2 = 0$. See Figure 2. In this case, we are able to obtain tight bounds on the flexibility of efficient rules that satisfy this condition (Section 4.1). Furthermore, even though this case yields a stronger condition than when $\epsilon_2 > 0$, there is a discontinuous increase in the range of such rules at $\epsilon_1 = 0$, demonstrating that a previous impossibility result is not robust to our weakening of *strategy-proofness.*

4.1 A Bound on the Range

Theorem 1 shows that the only efficient rules that are $(0,0)$-*strategy-proof* are those that always allocate all of the endowment to a prespecified agent. By relaxing the condition to $(\epsilon_1, 0)$-*strategy-proofness*, however, the class of admissible allocation rules is enlarged, as we show in Section 4.2.

We first provide, in this section, a result showing that in this case, such a rule always allocates nearly all of the endowment of good 1 to a prespecified agent. This can be considered a result corresponding to Theorem 1. However, as we show in the example of Section 4.2, the allocation of good 2 may vary across the entire range of feasibility. Hence Theorem 1 does not generalize in a continuous way.

Theorem 2 *Let φ be an efficient rule that is* $(\epsilon_1, 0)$-strategy-proof, *where $\epsilon_1 < \Omega^1/5$. There exists an agent $i \in N$ that always receives almost all of the good 1: for all $R \in \mathcal{R}^2$, $\varphi_i^1(R) \geq \Omega^1 - 2\epsilon_1$.*

To prove the result, we first provide the following lemma, which essentially states that for all preference profiles with the same set of efficient allocations, the chosen allocations of good 1 are not much different.

Lemma 1 *Let φ be efficient and* $(\epsilon_1, 0)$-strategy-proof. *For all $R, R' \in \mathcal{R}^2$, if either $E(R) = E(R') = E^{\ulcorner}$ or $E(R) = E(R') = E^{\lrcorner}$, then $|\varphi_1^1(R) - \varphi_1^1(R')| \leq 2\epsilon_1$.*

Proof: Let $R, R' \in \mathcal{R}^2$ be such that $E(R) = E(R') = E^{\ulcorner}$. It is either the case that $E(R_1, R_2') = E^{\ulcorner}$, or $E(R_1', R_2) = E^{\ulcorner}$. Without loss of generality, suppose $E(R_1, R_2') = E^{\ulcorner}$ (which is true, for example, if the indifference curves of R_1 are "flatter" than those of R_1').

By efficiency, $\varphi(R_1, R_2') \in E^{\ulcorner}$. Since φ is $(\epsilon_1, 0)$-*strategy-proof* and $\varphi(R_1', R_2') \in E^{\ulcorner}$, we have $\varphi_1^1(R_1, R_2') - \varphi_1^1(R_1', R_2') \leq \epsilon_1$. Similarly, $\varphi_1^1(R_1', R_2') - \varphi_1^1(R_1, R_2') \leq \epsilon_1$, so

$$|\varphi_1^1(R_1, R_2') - \varphi_1^1(R_1', R_2')| \leq \epsilon_1$$

By the same type of argument, we have

$$|\varphi_2^1(R_1, R_2) - \varphi_2^1(R_1, R_2')| \leq \epsilon_1$$

implying

$$|\varphi_1^1(R_1, R_2) - \varphi_1^1(R_1, R_2')| \leq \epsilon_1$$

Therefore, by the triangle inequality,

$$|\varphi_1^1(R_1, R_2) - \varphi_1^1(R_1', R_2')| \leq 2\epsilon_1$$

proving the result. □

Now we can prove the theorem.

Proof of Theorem 2: Let φ be efficient and $(\epsilon_1, 0)$-*strategy-proof*. There are three possible cases.

Case 1: For all $R \in \mathcal{R}^2$, if $E(R) = E^{\ulcorner}$, then $\varphi_1^2(R) = \Omega^2$.

Step 1a: ($E^{\lrcorner}$) In this case, for all $\delta > 0$, there exists $R \in \mathcal{R}^2$ such that $E(R) = E^{\lrcorner}$ and $\varphi_1(R) \geqq (\Omega^1, \Omega^2 - \delta)$. To see this, let R_1 satisfy $(0, \Omega^2)$ P_1 $(\Omega^1 + \epsilon_1, \Omega^2 - \delta)$, let R_2 be such that $E(R) = E^{\lrcorner}$, and let R_1' be such that $E(R_1', R_2) = E^{\ulcorner}$. Since φ is $(\epsilon_1, 0)$-*strategy-proof* and $E(R_1', R_2) = E^{\ulcorner}$,

$$\varphi_1(R) + (\epsilon_1, 0)\ R_1\ \varphi_1(R_1', R_2)\ R_1\ (0, \Omega^2)$$

by the hypothesis of Case 1. Therefore $\varphi_1(R)$ P_1 $(\Omega^1, \Omega^2 - \delta)$. Since $\varphi_1(R) \in E^{\lrcorner}$, we have $\varphi_1(R) \geqq (\Omega^1, \Omega^2 - \delta)$.

Therefore by Lemma 1, for all $R \in \mathcal{R}^2$, if $E(R) = E^{\lrcorner}$, then $\varphi_1^1(R) \geq \Omega^1 - 2\epsilon_1$.

Step 1b: ($E^{\ulcorner}$ and A) Let $R \in \mathcal{R}^2$ be such that $E(R) \in \{E^{\ulcorner}, A\}$, and suppose in contradiction to the theorem that $\Omega^1 - \varphi_1^1(R) - 2\epsilon_1 = \delta > 0$. Let $y, y', y'' \in E^{\lrcorner}$ satisfy (see Figure 3):

$$y_1\ I_1\ \varphi_1(R) + (\epsilon_1 + \tfrac{1}{3}\delta/3, 0)$$
$$y_1'\ I_1\ \varphi_1(R) + (\epsilon_1 + \tfrac{2}{3}\delta, 0)$$

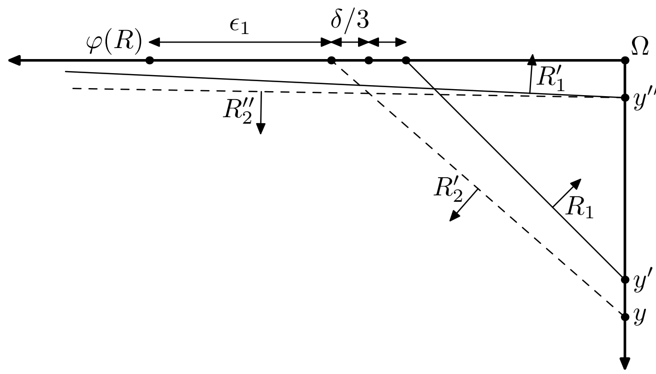


Figure 3: Proof of Theorem 2. The figure represents the upper-right corner of the Edgeworth Box.

$$y_1'' \ I_1 \ \varphi_1(R) + (2\epsilon_1 + \tfrac{2}{3}\delta, 0) = (\Omega^1 - \tfrac{1}{3}\delta, 0)$$

Let R_2' be such that $y_2 \ I_2' \ \varphi_2(R) - (\epsilon_1, 0)$. Since $\varphi(R_1, R_2') \in E^{\lrcorner}$, the truth-telling condition implies $\varphi_2(R_1, R_2') \geqq y_2$. Let R_2'' be sufficiently flat so that both $y_2'' \ P_2'' \ (\Omega^1 + \epsilon_1, 0)$ and $y_2 \ P_2'' \ y_2' + (\epsilon_1, 0)$. The truth-telling condition implies $\varphi_2(R_1, R_2'') + (\epsilon_1, 0) \ R_2'' \ \varphi_2(R_1, R_2')$, so $\varphi_2(R_1, R_2'') \geq y_2'$.

Let R_1' satisfy $(0, \Omega^2) \ I_1' \ y_1'' + (\epsilon_1, 0)$. Then $E(R_1', R_2'') = E^{\lrcorner}$. Note that by construction, $y_1' + (\epsilon_1, 0) \ I_1 \ y_1''$. The truth-telling condition implies $\varphi_1(R_1, R_2'') + (\epsilon_1, 0) \ R_1 \ \varphi_1(R_1', R_2'')$. Therefore $\varphi_1(R_1', R_2'') \leq y_1''$.

By the hypothesis of Case 1, for all R_1'' such that $E(R'') = E^{\ulcorner}$, we have $\varphi_1(R'') \geqq (0, \Omega^2)$. But then for any such R_1'', we have $\varphi_1(R'') \ P_1' \ \varphi_1(R_1', R_2'') + (\epsilon_1, 0)$, which contradicts the truth-telling condition.

Therefore, if Case 1 holds, we have derived the conclusion of the theorem.

Case 2: For all $R \in \mathcal{R}^2$, if $E(R) = E^{\lrcorner}$, then $\varphi_2^2(R) = \Omega^2$.

This case is symmetric to Case 1. In this case, for all $R \in \mathcal{R}^2$, $\varphi_2^1(R) \geq \Omega^1 - 2\epsilon_1$.

Case 3: Neither Case 1 nor Case 2 holds, i.e., there exist $R, R' \in \mathcal{R}^2$ such that $E(R) = E^{\ulcorner}$, $E(R') = E^{\lrcorner}$, $\varphi_1^2(R) < \Omega^2$, and $\varphi_2^2(R') < \Omega^2$.

In this case, by Lemma 1, for all $R, R' \in \mathcal{R}^2$, $E(R) = R^{\ulcorner}$ implies $\varphi_1^1(R) \leq 2\epsilon_1$, and $E(R') = R^{\lrcorner}$ implies $\varphi_2^1(R') \leq 2\epsilon_1$. Since $\epsilon_1 < \Omega^1/5$, this implies that

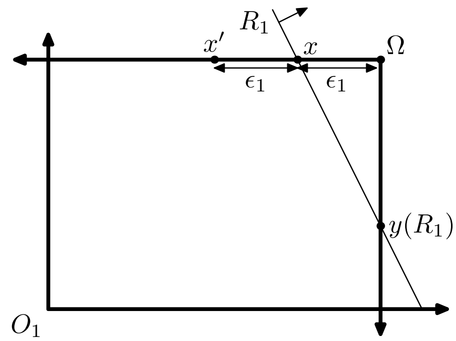


Figure 4: An efficient rule that is $(\epsilon_1, 0)$-*strategy-proof.*

for all such R, R',

$$\varphi_1^1(R') - \varphi_1^1(R) > \epsilon_1 \tag{1}$$

Let R_1 be such that $(2\epsilon_1, \Omega^2)\ P_1\ (\Omega^1 - 3\epsilon_1, 0)$. Let R_2, R_1' be such that $E(R) = E^{\ulcorner}$ and $E(R_1', R_2) = E^{\lrcorner}$. Then eqn. (1) implies $\varphi_1(R_1', R_2)\ P_1\ \varphi_1(R) + (\epsilon_1, 0)$, which contradicts the truth-telling condition. Therefore this case cannot hold. □

4.2 An Important Rule

Theorem 2 states that under an efficient rule that is $(\epsilon_1, 0)$-*strategy-proof*, one agent always must receive at least $\Omega^1 - 2\epsilon_1$ of good 1. The rule described below in Example 1 shows that (i) this bound is tight, and (ii) there is no such bound corresponding to good 2. In the example, agent 1 receives (i) from as little as $\Omega^1 - 2\epsilon_1$ of the good 1 to as much as all of it, and (ii) from as little as none of good 2 to as much as all of it.

Furthermore, and *most importantly*, we provide Theorem 3, showing that this rule is, unambiguously, the "least dictatorial" (or most equitable) of all efficient rules that are $(\epsilon_1, 0)$-*strategy-proof.* This result is not provided as a justification for the use of the given rule. Instead, it is used later to show a discontinuity when we relax the condition to (ϵ_1, ϵ_2)-*strategy-proofness* for $\epsilon_2 > 0$.

Example 1 Fix the allocations $x = ((\Omega^1 - \epsilon_1, \Omega^2), (\epsilon_1, 0))$, which gives agent 1 the entire endowment except for ϵ_1 units of good 1, and $x' = ((\Omega^1 - 2\epsilon_1, \Omega^2), (2\epsilon_1, 0))$. For all $R_1 \in \mathcal{R}$, let $y(R_1) \in E^{\lrcorner}$ be the unique allocation in $E^{\lrcorner}$ that agent 1 considers indifferently to x (as in Figure 4), i.e., that $x_1 \; I_1 \; y_1(R_1)$. Then for all $R \in \mathcal{R}^2$, let

$$\tilde{\varphi}(R) = \begin{cases} x' & \text{if } x' \text{ is efficient for } R \\ y(R_1) & \text{otherwise} \end{cases}$$

We leave it to the reader to check that $\tilde{\varphi}$ is efficient and $(\epsilon_1, 0)$-*strategy-proof*.[10]

This rule is clearly not symmetric. In fact, for most profiles of preferences, both agents would prefer agent 1's consumption bundle to agent 2's. A more formal welfare analysis appears in Section 6. The statement of Theorem 2 does not, by itself, rule out more equitable rules. As the next theorem shows, however, $\tilde{\varphi}$ is the most equitable efficient rule that is $(\epsilon_1, 0)$-*strategy-proof*. Under any other such rule, say φ, one of the two agents would, *under any profile of preferences*, prefer the bundle that $\tilde{\varphi}$ prescribes to agent 2 to the one that φ prescribes to that agent.

THEOREM 3 *No efficient,* $(\epsilon_1, 0)$-strategy-proof *rule is more equitable than* $\tilde{\varphi}$.[11] *Specifically, let* φ *be an efficient rule that is* $(\epsilon_1, 0)$-strategy-proof, *where* $\epsilon_1 < \Omega^1/5$. *Then one of the agents must (weakly) prefer playing the role of agent 2 under* $\tilde{\varphi}$ *to playing his own role under* φ*, i.e., one of the following is true.*

(1) for all $R \in \mathcal{R}^2$, $\tilde{\varphi}_2(R) \; R_2 \; \varphi_2(R)$ *(agent 2 prefers* $\tilde{\varphi}$ *to* φ*), or,*

(2) for all $R \in \mathcal{R}^2$, $\tilde{\varphi}_2(R') \; R_1 \; \varphi_1(R)$, *where* $R'_1 = R_2$ *and* $R'_2 = R_1$ *(agent 1 prefers playing the role of agent 2 under* $\tilde{\varphi}$ *to what he gets under* φ*).*

Proof: Let φ be an efficient rule that is $(\epsilon_1, 0)$-*strategy-proof*. Suppose (by Theorem 2) that agent 1 is the agent who always receives at least $\Omega^1 - 2\epsilon_1$

[10]Clearly, a mirror image to this rule exists in which the labels of the two agents are switched, and that rule also satisfies the two properties.

[11]The same result obviously applies to the rule obtained from $\tilde{\varphi}$ by reversing the (asymmetric) roles of the two agents.

of the numeraire good under φ. In this case, we show statement (1) of the Theorem: for all $R \in \mathcal{R}^2$, $\tilde{\varphi}_2(R)\ R_2\ \varphi_2(R)$. (If we were to suppose instead that agent 2 always does, then we would prove statement (2).)

If $E(R) = E^{\ulcorner}$, the conclusion follows from Theorem 2, since $\varphi_2(R) \leqq (2\epsilon_1, 0) = \tilde{\varphi}_2(R)$.

If either $E(R) = E^{\lrcorner}$ or $E(R) = A$, suppose in contradiction to the theorem that $\varphi_2(R)\ P_2\ \tilde{\varphi}_2(R)$. Then there exists $\delta > 0$ such that

$$\varphi_1(R)\ I_1\ (\Omega^1 - \epsilon_1 - \frac{2}{3}\delta, \Omega^2)$$

(otherwise the proof is trivial). Letting $y = \varphi(R)$ and $R_2' = R_2$, and defining y', y'', R_1', and R_1'' as in Figure 3 (proof of Theorem 2), leads to a contradiction as it did in Step 1b of that proof. □

In light of this result, a full characterization of efficient, $(\epsilon_1, 0)$-*strategy-proof* rules does not appear to be interesting. For example, the rule $\tilde{\varphi}$ can be perturbed in many uninteresting ways (e.g., by giving slightly more of the goods to agent 1) while remaining $(\epsilon_1, 0)$-*strategy-proof*.

4.3 A Discontinuity at $\epsilon_1 = 0$

To emphasize the idea that a small relaxation in *strategy-proofness* leads to a large increase in the flexibility of rules, consider the implications of $(\epsilon_1, 0)$-*strategy-proofness* as ϵ_1 approaches zero. The rule $\tilde{\varphi}$ was defined in Example 1 with respect to a given value of ϵ_1. The range of this rule for a given ϵ_1 is

$$\{x \in E^{\lrcorner} : x_1^1 > \Omega^1 - \epsilon_1,\ x_1^2 < \Omega^2\} \cup \{((\Omega^1 - 2\epsilon_1, \Omega^2), (2\epsilon, 0))\}$$

As ϵ_1 converges to zero, this set converges to the right-hand border of the Edgeworth Box, i.e., to $\{x \in A : x_1^1 = \Omega^1\}$.

Therefore, as $(\epsilon_1, 0)$-*strategy-proofness* converges to *strategy-proofness*, the range of admissible rules does *not* converge to the class of *strategy-proof* and efficient rules (i.e., dictatorial rules) characterized in Schummer (1997).[12]

[12] Formally, this sequence of examples shows that the ranges of the admissible rules is a correspondence that is not upper-semi-continuous at $\epsilon_1 = 0$, fixing $\epsilon_2 = 0$. It is clearly lower-semi-continuous: dictatorial rules are $(\epsilon_1, 0)$-*strategy-proof* for any ϵ_1.

This discontinuity is important to observe because it reinforces the notion that a small relaxation of *strategy-proofness* leads to a relatively large increase in the range of admissible rules. On domains for which impossibility results regarding *strategy-proofness* have been established, relaxing the condition even in a small way may significantly enlarge the class of admissible allocation rules.

5 Rules for (ϵ_1, ϵ_2)-*strategy-proofness*

We now turn our attention to the weaker condition of (ϵ_1, ϵ_2)-*strategy-proofness* when $\epsilon_2 > 0$. The rule $\tilde{\varphi}$ can be generalized in various ways, not all of which are obvious. For example, an obvious generalization could be obtained in part by redefining x to be $(\Omega_1 - \epsilon_1, \Omega_2 - \epsilon_2)$, and generalizing the rule in the obvious way.

A generalization that is less asymmetric (and perhaps less obvious) is defined in the following example.

Example 2 For all $R_1 \in \mathcal{R}$, let $y(R_1) \in E^{\lrcorner}$ and $z(R_1) \in E^{\ulcorner}$ be the unique allocations (as in Figure 5) such that $y_1(R_1)\ I_1\ (\Omega_1 - \epsilon_1, \Omega_2 - 2\epsilon_2)$ and $z_1(R_1)\ I_1\ (\Omega_1 - 2\epsilon_1, \Omega_2 - \epsilon_2)$. Then for all $R \in \mathcal{R}^2$, let

$$\hat{\varphi}(R) = \begin{cases} z(R_1) & \text{if } z(R_1) \text{ is efficient for } R \\ y(R_1) & \text{otherwise} \end{cases}$$

We leave it to the reader to check that $\hat{\varphi}$ is efficient and (ϵ_1, ϵ_2)-*strategy-proof*.

5.1 Discontinuity at $\epsilon_2 = 0$ for any ϵ_1

We showed in section 4.3 that relaxing *strategy-proofness* to $(\epsilon_1, 0)$-*strategy-proofness* results in a discontinuous increase in the range of admissible, efficient rules. That discontinuity holds as ϵ_1 approaches zero, holding $\epsilon_2 = 0$ fixed.

The rule $\hat{\varphi}$ shows that further relaxing to (ϵ_1, ϵ_2)-*strategy-proofness* with $\epsilon_2 > 0$ results in a more striking discontinuity: for any $\epsilon_1 \geq 0$, the range of admissible rules (as a function of ϵ_2) is discontinuous at $\epsilon_2 = 0$.

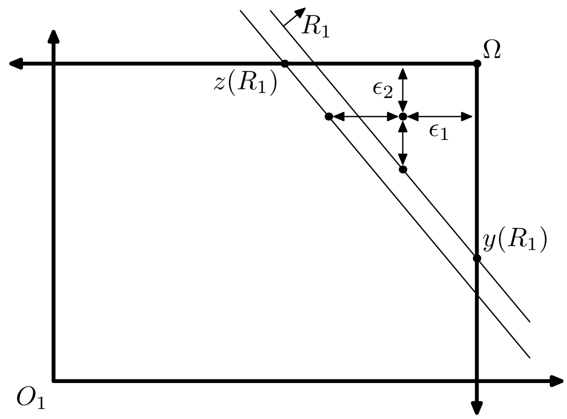


Figure 5: Defining an efficient rule that is (ϵ_1, ϵ_2)-*strategy-proof*.

To see this, observe that the range of $\hat{\varphi}$ (defined with respect to a given $\epsilon > (0, 0)$) is

$$\{x \in E^{\lrcorner} : x_1^1 > \Omega^1 - \epsilon_1,\ x_1^2 < \Omega^2 - 2\epsilon_2\} \cup \{x \in E^{\ulcorner} : x_1^2 > \Omega^2 - \epsilon_2,\ x_1^1 < \Omega^1 - 2\epsilon_1\}$$

As $\epsilon > (0, 0)$ converges to $(0, 0)$, this range converges to the upper and right-hand borders of the Edgeworth Box, i.e., to

$$\{x \in A : x_1^1 = \Omega^1 \text{ or } x_1^2 = \Omega^2\}$$

However, by Theorem 2, the range of any efficient, $(\epsilon_1, 0)$-*strategy-proof* rule is contained in the set $\{x \in A : x_1^1 \geq \Omega^1 - 2\epsilon_1\}$, which converges to $\{x \in A : x_1^1 = \Omega^1\}$ as ϵ_1 converges to zero.

To summarize, see Figure 6, showing the ranges (a) of $\tilde{\varphi}$ and (b) of $\hat{\varphi}$. Recall that an efficient, *strategy-proof* rule has a range consisting of one point: either the point labelled Ω or the point labelled O_1. Sending $\epsilon_1 \to 0$, Figure 6a shows a discontinuity between such a dictatorial range and $\tilde{\varphi}$, and hence shows discontinuity between the consequences of *strategy-proofness* and of $(\epsilon_1, 0)$-*strategy-proofness*. Furthermore, Theorem 2 shows that no other sequence of $(\epsilon_1, 0)$-*strategy-proof* rules (for $\epsilon_1 \to 0$) can converge to a set larger than the limit of (a) as ϵ_1 converges to 0. Therefore, sending $(\epsilon_1, \epsilon_2) \to (0, 0)$, Figure 6b shows a discontinuity not only between (a) and (b), but between $(\epsilon_1, 0)$-*strategy-proofness* and (ϵ_1, ϵ_2)-*strategy-proofness* for

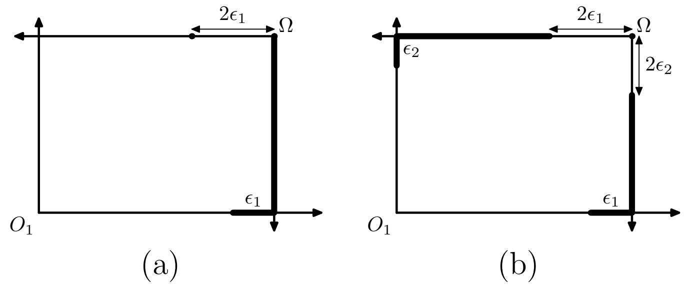


Figure 6: The ranges of (a) $\tilde{\varphi}$ and (b) $\hat{\varphi}$.

$\epsilon_2 > 0$.

6 Measures of Welfare

First consider the stronger condition of $(\epsilon_1, 0)$-*strategy-proofness*. Theorem 3 provides a lower bound on the welfare of the "unfavored" agent under an efficient, $(\epsilon_1, 0)$-*strategy-proof* rule. In order to have a better understanding of how well-off agent 2 is under the rule $\tilde{\varphi}$, it is useful to consider a class of normalized utility functions. We parameterize each preference relation $R_i \in \mathcal{R}$ with $\lambda_i \in]0, 1[$ such that the preference relation is represented by the utility function

$$u(x_i) = \lambda_i x_i^1 + (1 - \lambda_i) x_i^2$$

Below we consider the case in which $\Omega = (1, 1)$. In this case, an agent's utility is always equal to one when he receives the entire endowment, and is always equal to zero when he receives nothing. In particular, a utility level can be interpreted as a proportion of the entire endowment, that is, $u(\delta, \delta) = \delta$.

Under the rule $\tilde{\varphi}$ (defined with respect to a given ϵ_1), agent 2's utility is a function of λ_1, λ_2, ϵ_1, and Ω. It is a fairly straightforward geometric exercise to derive agent 2's utility under $\tilde{\varphi}$:[13]

$$u_2(\tilde{\varphi}; \lambda, \epsilon, \Omega) = \begin{cases} 2\lambda_2\epsilon_1 & \text{if } \lambda_2 \geq \lambda_1; \\ (1 - \lambda_2)\epsilon_1\lambda_1/(1 - \lambda_1) & \text{if } \lambda_2 < \lambda_1 \leq \Omega^2/(\Omega^2 + \epsilon_1); \\ \lambda_2\epsilon_1 + \Omega^2(\lambda_1 - \lambda_2)/\lambda_1 & \text{otherwise.} \end{cases}$$

[13]A proof is available upon request.

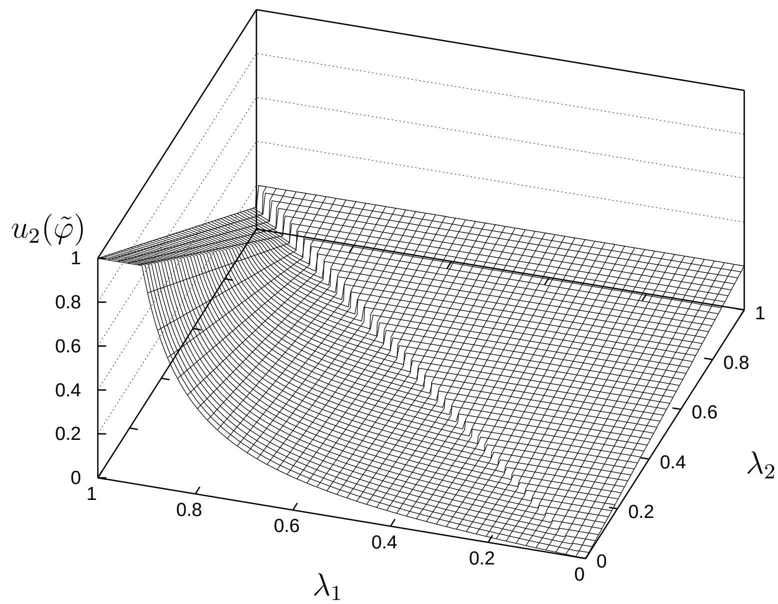


Figure 7: Utility to agent 2 from the rule $\tilde{\varphi}$, when $\epsilon_1 = 0.1$. Higher λ_i indicates higher relative preference toward good 1.

Figure 7 graphs $u_2(\tilde{\varphi};\cdot)$ when $\epsilon_1 = 0.1$ and $\Omega = (1,1)$. We see that agent 2 receives a non-negligible amount of utility at most profiles. The average utility that agent 2 receives over this entire range of values for (λ_1, λ_2) is approximately 0.18 (under a uniform distribution).[14] This is significantly higher than the average utility agent 2 would receive by consuming a constant $\epsilon_1 = 0.1$ units of good 1, which would be 0.05 units of utility.

By the previously mentioned result of Schummer (1997), if *strategy-proofness* were required, one of the agents would receive a constant utility of zero. These numbers encourage the idea that a "small" relaxation of *strategy-proofness* leads in some sense to a "larger" relaxation of dictatorship.

When the condition is weakened to (ϵ_1, ϵ_2)-*strategy-proofness* with $\epsilon_2 > 0$, the rule $\hat{\varphi}$ is admissible. Under that rule (defined with respect to a given ϵ),

[14]Upon request, an Excel file is available to compute $u_2(\tilde{\varphi};\cdot)$.

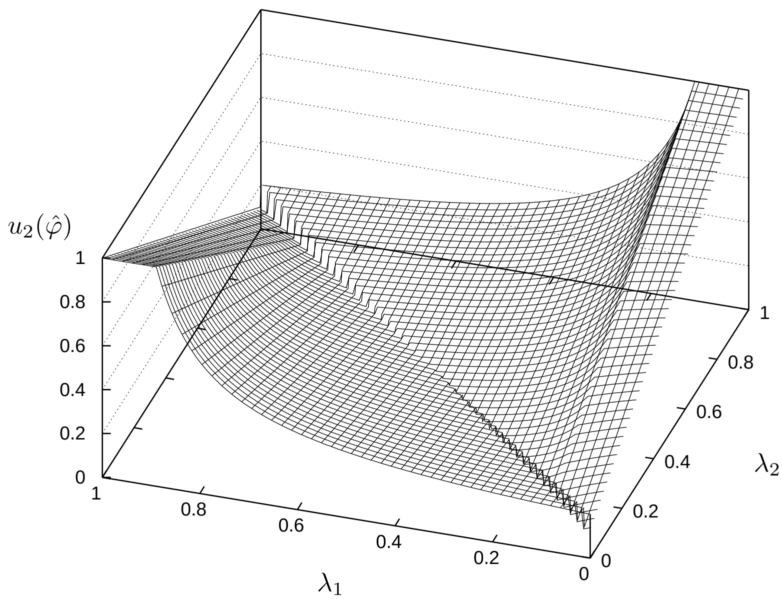


Figure 8: Utility to agent 2 from the rule $\hat{\varphi}$, when $\epsilon_1 = \epsilon_2 = 0.1$.

agent 2's utility is

$$u_2(\hat{\varphi};\lambda,\epsilon,\Omega) = \begin{cases} \lambda_2\Omega^1 + (1-\lambda_2)(\epsilon_2 + (2\epsilon_1 - \Omega^1)\lambda_1/(1-\lambda_1) \\ \qquad\qquad \text{if } \lambda_2 \geq \lambda_1 \text{ and } 2\epsilon_1 + \epsilon_2(1-\lambda_1)/\lambda_1 > \Omega^1; \\ \lambda_2(2\epsilon_1 + \epsilon_2(1-\lambda_1)/\lambda_1) \\ \qquad\qquad \text{if } \lambda_2 \geq \lambda_1 \text{ and } 2\epsilon_1 + \epsilon_2(1-\lambda_1)/\lambda_1 < \Omega^1; \\ (1-\lambda_2)\Omega^2 + \lambda_2(\epsilon_1 + (2\epsilon_2 - \Omega^2)(1-\lambda_1)/\lambda_1) \\ \qquad\qquad \text{if } \lambda_2 < \lambda_1 \text{ and } 2\epsilon_2 + \epsilon_1\lambda_1/(1-\lambda_1) > \Omega^2; \\ (1-\lambda_2)(2\epsilon_2 + \epsilon_1\lambda_1/(1-\lambda_1)) \\ \qquad\qquad \text{if } \lambda_2 < \lambda_1 \text{ and } 2\epsilon_2 + \epsilon_1\lambda_1/(1-\lambda_1) < \Omega^2. \end{cases}$$

Figure 8 graphs agent 2's utility under this rule when $\epsilon_1 = \epsilon_2 = .1$ and $\Omega = (1,1)$. The average value is approximately .34.[15]

[15] Upon request, an Excel file is available to compute $u_2(\hat{\varphi};\cdot)$.

7 Other Domains

So far we have restricted attention to exchange economies with only two goods. There is an obvious generalization of our definition of (ϵ_1, ϵ_2)-*strategy-proofness* to settings with more than two goods. Furthermore, the definition of (ϵ_1, ϵ_2)-*strategy-proofness* applies, as given, to more general classes of preferences, such as the standard domain of quasi-linear, monotonic preferences. In this section, we briefly mention such generalizations of the above analysis.

For economies with k goods (and linear preferences), the natural definition of our condition is as follows.

***ϵ-strategy-proofness*:** For any $\epsilon \in \mathbb{R}^k_+$, a rule is ϵ-*strategy-proof* if for all $R \in \mathcal{R}^2$, all $i \in \{1, 2\}$, and all $R'_i \in \mathcal{R}$, we have

(j) $\varphi_i(R) + (0, \ldots, 0, \epsilon_j, 0 \ldots, 0)\ R_i\ \varphi_i(R'_i, R_{-i})$

for some j, $1 \leq j \leq k$.

The rules $\tilde{\varphi}$ and $\hat{\varphi}$ can be generalized in a natural way. Formally, generalize $\tilde{\varphi}$ by letting x, x' be the allocations such that $x_1 = \Omega - (\epsilon_1, 0, \ldots, 0)$ and $x'_1 = \Omega - (2\epsilon_1, 0, \ldots, 0)$; let $y(R_1, R_2)$ be the set of efficient allocations that agent 1 considers indifferently to x. Then, $\tilde{\varphi}$ (defined as before) is efficient and $(\epsilon_1, 0, \ldots, 0)$-*strategy-proof.*

Furthermore, it seems clear that for this case, results analogous to Theorems 2 and 3 can be obtained, showing this generalization of $\tilde{\varphi}$ to be the "least dictatorial" such rule on this domain. With an investment in additional notation, such results should be obtainable in the same way Schummer (1997) extends the results for 2-agent/2-good economies to multiple-good economies. For brevity, we omit this notationally tedious task.

In a similar way, $\tilde{\varphi}$ can also be generalized to the domain of economies in which two agents may have any (possibly non-linear) quasi-linear preference relation.[16] In this case, generalize the notation from Example 1 (e.g., for the 2-good case) by letting $y(R_1, R_2)$ be any efficient allocation that agent 1 considers indifferently to x, and letting $x'(R_1, R_2)$ (now a function) be any efficient allocation that agent 1 considers indifferently to $\Omega - (2\epsilon_1, 0)$.

[16] We omit a formalization of this standard domain of preferences.

8 Lotteries

The ideas behind the construction of $\tilde{\varphi}$ and $\hat{\varphi}$ can be extended to a voting environment in which agents have cardinal preferences over outcomes, and a rule chooses a lottery over outcomes. In this section, we examine such an environment.

The problem of choosing lotteries over public outcomes has a flavor slightly different than the previous problem, which involved private goods. Here, there are situations in which agents agree on what is the best outcome, in which case efficiency requires the choice of that outcome.

8.1 The Lotteries Model

For simplicity, we examine the case in which there are exactly three pubic outcomes, $\{X, Y, Z\}$. The analysis extends in an obvious way to the case of more outcomes. A *lottery* is a triple $\ell = (\ell_x, \ell_y, \ell_z) \in \mathbb{R}^3_+$ such that $\ell_x + \ell_y + \ell_z = 1$. Let $\triangle$ denote the set of lotteries. Each of the two agents has a von Neumann-Morgenstern preference relation, $\succsim_i$, over lotteries, i.e., a preference relation representable by a utility function of the form $u(\ell) = \lambda_x \ell_x + \lambda_y \ell_y + \lambda_z \ell_z$, where $\lambda_x, \lambda_y, \lambda_z \in \mathbb{R}$. Let $\mathcal{R}_L$ denote the set of such preference relations.[17]

A rule $\varphi\colon \mathcal{R}^2_L \to \triangle$ assigns a lottery $\varphi(\succsim)$ to every pair of preference relations $\succsim = (\succsim_1, \succsim_2)$. A rule is (ex-ante) efficient if it always chooses Pareto optimal lotteries, i.e., there never exists a lottery that both agents prefer to the one chosen by the rule, with one agent strictly preferring it.

In this environment, we weaken *strategy-proofness* by requiring only that any lottery obtained through a manipulation cannot be preferred (by the manipulating agent) to every lottery within an ϵ-ball of the original one; see Figure 9. Formally, for $\epsilon \in \mathbb{R}_+$, a rule φ is *ϵ-strategy-proof* if for all $i \in \{1, 2\}$ and all $\succsim_1, \succsim_2, \succsim'_i \in \mathcal{R}_L$, there exists $(\ell'_x, \ell'_y, \ell'_z) \in \triangle$ such that both $||\ell' - \varphi(\succsim)|| \leq \epsilon$, and $\ell' \succsim_i \varphi(\succsim'_i, \succsim_{-i})$.

[17] In what follows, we ignore the "total indifference" relation, where $\lambda_x = \lambda_y = \lambda_z$. Including it changes no results, but adds to the notation.

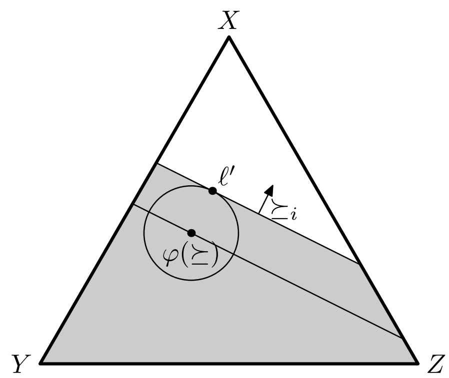


Figure 9: Definition of ϵ-*strategy-proofness*. Lotteries attainable by agent i must lie in the shaded area.

8.2 A Rule

The ideas behind the rules introduced for exchange economies can be applied to the lotteries model. Again, a discontinuity in the range of admissible rules occurs as *strategy-proofness* is weakened to ϵ-*strategy-proofness* for $\epsilon > 0$. Hylland (1980)[18] shows that the only efficient rule that is 0-*strategy-proof* is dictatorial: a rule simply chooses a prespecified agent's favorite outcome with probability one. When *strategy-proofness* is weakened, however, we have (among others) the following class of rules.

Example 3 Let $0 < \gamma < 1/3$. We define here the rule φ^γ for the case in which agent 1 strictly prefers X over the other two alternatives. The rule is defined symmetrically for the other cases simply by relabelling the outcomes.

Fix the lottery $\ell^X = (1-2\gamma, \gamma, \gamma)$. We now define $y(\succeq_1)$ and $z(\succeq_1)$ as in Fig. 10. For any $\succsim_1 \in \mathcal{R}_L$ such that $X \succsim_1 Y$ and $X \succsim_1 Z$, let $y'(\succsim_1) \in \triangle$ be the unique lottery such that (i) $y'(\succsim_1) \sim_1 \ell^X$, (ii) $\min\{y'_x(\succsim_1), y'_z(\succsim_1)\} = 0$, and (iii) $y'_z(\succsim_1) \leq \gamma$. Let $y(\succsim_1) \in \triangle$ be such that (i) if $y'_x(\succsim_1) \geq \gamma$, then $y(\succsim_1) = y'(\succsim_1)$, and (ii) otherwise, $y(\succsim_1) = (\gamma, 1-\gamma, 0)$. Define $z'(\succsim_1)$ and $z(\succsim_1)$ symmetrically.

[18]See also, Gibbard (1977) and Schummer (1999).

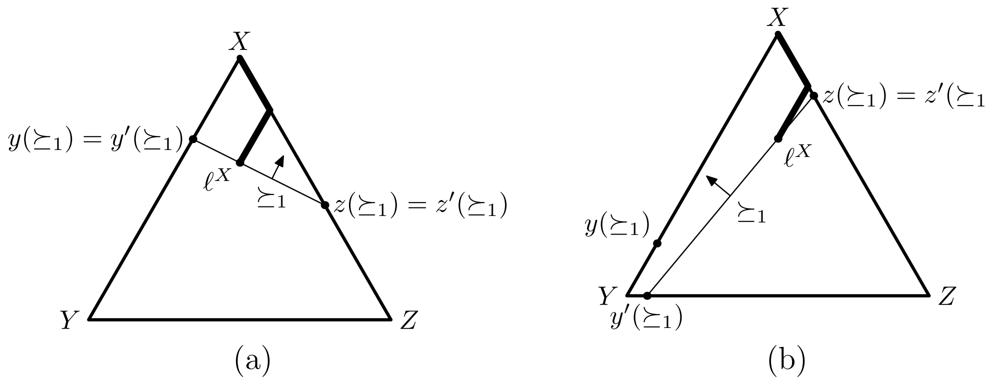


Figure 10: Definition of an ϵ-*strategy-proof* rule for the case $X \succsim_1 Y \succsim_1 Z$. Each dark line has length $\gamma\sqrt{2}$, where the length of each side of the simplex is $\sqrt{2}$. Figure (b) illustrates an instance where $y \neq y'$.

For the case in which $X \succ_1 Y$ and $X \succ_1 Z$, the rule φ^γ is defined as

$$\varphi^\gamma(\succsim) = \begin{cases} y(\succsim_1) & \text{if } y(\succsim_1) \text{ is efficient for } \succsim, \\ z(\succsim_1) & \text{if } z(\succsim_1) \text{ is efficient and } y(\succsim_1) \text{ is not,} \\ X & \text{otherwise} \end{cases}$$

For the cases in which $Y \succsim_1 X$ or $Z \succsim_1 X$, the definition of φ^γ is obtained as above by relabelling the outcomes.

The rule is clearly efficient by definition. Additionally, when γ is small enough, the rule is almost *strategy-proof* in the sense defined above.

Proposition 1 *When* $\gamma \leq \epsilon$, φ^γ *is* ϵ-strategy-proof.

Proof: (Sketch) Agent 2 can never manipulate the rule at all; it is 0-*strategy-proof* from his perspective.

Consider the case in which Agent 1 considers X the best outcome. The rule either chooses X (when both agents agree it is best), or chooses a lottery he considers indifferently to ℓ^X. When $\gamma \leq \epsilon$, the only outcome in the range of φ^γ that could be more than "ϵ-better" than ℓ^X is X itself. By the definition of the rule, however, that outcome is only attainable when agent 2 considers X to be best. □

8.3 Welfare analysis

In this section, we examine the welfare effect of relaxing *strategy-proofness* to *ϵ-strategy-proofness*. As mentioned previously, the only (efficient) rules that are *strategy-proof* are dictatorial: they always choose a prespecified individual's favorite outcome (with probability one). Below, we compare the utility agent 2 would receive under the rule in which agent 1 is a "dictator" to the utility he would receive under φ^γ.

Since φ^γ is defined symmetrically with respect to the outcomes $\{X, Y, Z\}$, we can restrict attention to the case in which agent 1 has preferences such that $X \succ_1 Y \succ_1 Z$, while varying agent 2's preferences over the entire domain. The other five cases for agent 1's preferences result in analyses that are symmetric to the analysis below.

We use the following utility representation of preferences. For any preference relation, the "best" outcome is assigned a utility of 1, the "worst" is assigned utility of 0, and the middle outcome is assigned a utility of $\lambda_i \in [0, 1]$. Since we fix agent 1's ordinal ranking of the three outcomes (and ignore the case of indifference), his preferences are determined by the parameter $\lambda_1 \in]0, 1[$, so that $Y \sim_1 (\lambda_1, 0, 1 - \lambda_1)$.

Agent 2's preferences are described by both a ranking of the outcomes $\{X, Y, Z\}$ and the parameter $\lambda_2 \in [0, 1]$. For example, if $Y \succ_2 Z \succ_2 X$, then we are using the utility representation $u(x, y, z) = 0x + 1y + \lambda_2 z$ for the appropriate λ_2.

For the case $\gamma = .10$, Figure 11 shows Agent 2's utility as a function of (i) his preferences (an ordering and a value for λ_2) and (ii) λ_1 (fixing $X \succ_1 Y \succ_1 Z$).[19] Roughly speaking, the axis for agent 2 provides the angle of his indifference lines in the simplex, over 360 degrees of rotation.

For example, when both agents agree that X is the best outcome (i.e., in the figure, the first and sixth orderings for agent 2), both agents receive maximal utility (equal to one). Note, however, that this would be true under *any* rule that is efficient. Therefore, it does not make sense to simply

[19]To completely graph Agent 2's utility over the entire domain of preferences, we would construct five more versions of Figure 11, for each other ordering for agent 1. Each such graph would appear as a "permutation" of the six regions in Figure 11.

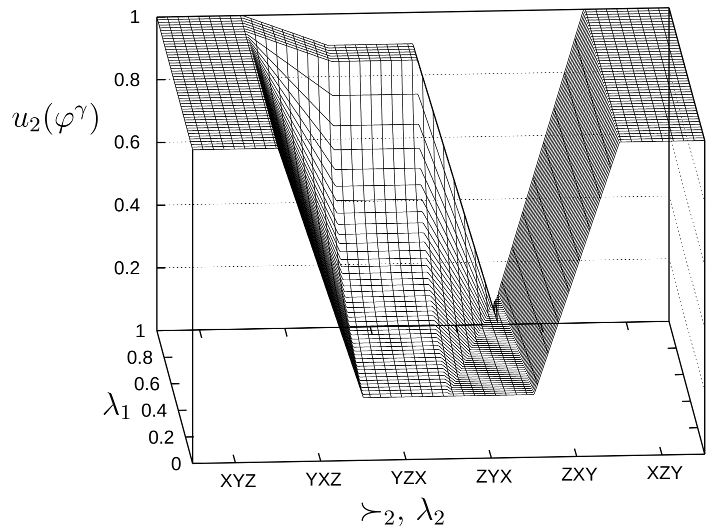


Figure 11: Agent 2's utility under φ^γ when $\gamma = .10$.

examine the utility to Agent 2 under φ^γ. It is more insightful to examine the utility *gains* to Agent 2 from using φ^γ instead of using the rule that always chooses Agent 1's favorite outcome. This tells us what is gained from relaxing *strategy-proofness* to ϵ-*strategy-proofness*.

Figure 12 shows the additional utility Agent 2 receives when the rule φ^γ is used instead of the rule that always chooses Agent 1's most-preferred outcome, again for the case $\gamma = .10$.

What is interesting to observe is that when agent 1 is "almost" indifferent between X and Y, we see a significant increase in agent 2's utility by using our ϵ-*strategy-proof* rule rather than the dictatorial rule. This parallels the analysis of exchange economies. There, agent 2's gains are greatest in situations where agent 1 places relatively little value on one good, and agent 2 places relatively little value on the other good.

Numerically, we have the following approximations of average utility (under uniform independent distributions on the λ_i's, for the cases in which $X \succsim_1 Y \succsim_1 Z$), given agent 2's ordering of the three alternatives. The first

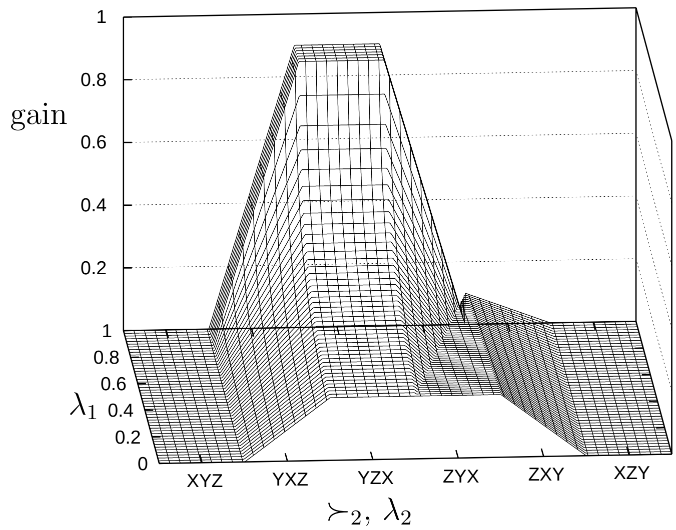


Figure 12: The utility gains to agent 2 from using φ^γ ($\gamma = .10$) instead of the agent-1-dictatorial rule.

column gives the ordering; the second gives agent 2's average utility under φ^γ, given that ordering; the third gives his average utility, given that ordering, under the rule in which agent 1 is a dictator. Each of the six cases corresponds to one-sixth of Figure 11.

ordering of $\succsim_2$	2's utility (φ^γ)	2's utility (dictatorial rule)
$X \succsim_2 Y \succsim_2 Z$	1	1
$Y \succsim_2 X \succsim_2 Z$	.71	0.5
$Y \succsim_2 Z \succsim_2 X$	.41	0
$Z \succsim_2 Y \succsim_2 X$	.25	0
$Z \succsim_2 X \succsim_2 Y$	.58	0.5
$X \succsim_2 Z \succsim_2 Y$	1	1

The largest average utility gains over the (agent 1) dictatorial rule occur when $Y \succsim_2 Z \succsim_2 X$; see Figure 12.

9 Conclusion

We have introduced a relaxation of the notion of dominant strategy implementation by requiring only that truth-telling be an almost-dominant strategy. This concept is first formalized in an exchange economy model by defining ϵ-*strategy-proofness*: we only eliminate misrepresentations that would give an agent a gain better than receiving ϵ_j of any good j. A similar formalization is made in a voting model with randomization, where comparisons are made in lottery-space.

On the simple class of 2-agent exchange economies with two goods, in which agents have linear preference relations, we have provided (Theorem 2) a bound on the range of any efficient rule that is $(\epsilon_1, 0)$-*strategy-proof*: A prespecified agent must always receive almost all of good 1; the second agent always receives at most $2\epsilon_1$ units of the good. However, we provide a rule (Example 1) which varies the allocation of the second good between the two agents to the degree that in some situations, one agent receives all of it, while in some other situations, the other agent receives all of it.

The flexibility of this rule is in stark contrast to the conclusions derived when truth-telling is required to be a fully dominant strategy, i.e., that the only *strategy-proof*, efficient rule in this context always allocates *all* of the goods to a prespecified agent (Schummer (1997)). Admittedly, the rule provided in Example 1 is not flexible in its allocation of the first good. However, we show (in Theorem 3) that this rule is actually the most equitable of all efficient rules that are $(\epsilon_1, 0)$-*strategy-proof*. Furthermore, the rule shows that the class of admissible rules is discontinuous at $\epsilon_1 = 0$.

When *strategy-proofness* is weakened even further to (ϵ_1, ϵ_2)-*strategy-proofness* ($\epsilon_2 > 0$), the class of admissible rules discontinuously increases again. A rule exists which is flexible in its allocation of *both* goods.

A similar discontinuity occurs in a 2-agent voting model with lotteries. Previous results showed that the only (ex-ante) efficient, *strategy-proof* rules in this model are dictatorial. However, as we weaken the condition to (ϵ_1, ϵ_2)-*strategy-proofness*, we again see a discontinuous increase in the range of admissible rules.

The negative interpretation of the results is that (ϵ_1, ϵ_2)-*strategy-proofness* does restrict our choice of rules in these environments. This is not surprising, given the previous results concerning *strategy-proofness*.

The positive interpretation of the results is the following: Even a small relaxation of *strategy-proofness* leads to a relatively large increase in the flexibility of rules. This is good news not only for domains with previously established impossibility results regarding *strategy-proofness*, but also for other domains in which additional requirements—such as efficiency—may restrict our choice of reasonable allocation rules.

There is an additional point that gives these results even more positive flavor. In models with additional agents, the rules satisfying the truth-telling condition may be even more flexible. The $2\epsilon_1$-bound of Theorem 1 crucially depends on the fact that there are only two agents. Roughly speaking, two unilateral changes in preferences change the welfare of agents by an amount comparable to at most $2\epsilon_1$ units of good 1 (as in the proof of Theorem 2). With more agents, there is reason to believe that changes in preferences by more agents will lead to even greater flexibility in rules satisfying our condition.[20] This provides hope that for other economic environments for which impossibility results have been obtained regarding *strategy-proofness*, there is good reason for a mechanism designer to consider our weaker truth-telling condition as a second-best incentives constraint.

References

Abreu, D. and H. Matsushima (1992). "Virtual Implementability in Iteratively Undominated Strategies: Complete Information," *Econometrica*, **60**, 993–1008.

Alcaldé, J. and S. Barberà (1994). "Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems," *Economic Theory*, **4**, 417–35.

Barberà, S. (1977). "The Manipulation of Social Choice Mechanisms that Do Not Leave 'Too Much' to Chance," *Econometrica*, **45**, 1573–88.

[20]This could be seen as a dual result to the asymptotic results of Roberts and Postlewaite (1976).

Barberà, S. (2001). "An Introduction to Strategy-proof Social Choice Functions," *Social Choice and Welfare*, **18**, 619–653.

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

Barberà, S. and B. Peleg (1990). "Strategy-proof Voting Schemes with Continuous Preferences," *Soc Choice Welfare*, **7**, 31–8.

Beviá, C. and Corchón, L. (1995). "On the Generic Impossibility of Truthful Behavior: A Simple Approach," *Econ Theory* **6**, 365–71.

Ching, S. (1994). "An Alternative Characterization of the Uniform Rule," *Social Choice and Welfare*, **11**, 131-36.

Ching, S. (1992). "A Simple Characterization of the Uniform Rule," *Economics Letters*, **40**, 57–60

Ching, S. (1997). "Strategy-Proofness and 'Median Voters'," *International Journal of Game Theory*, **26**, 473–90.

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

Córdoba, J. and P. Hammond (1998). "Asymptotically Strategy-Proof Walrasian Exchange," *Mathematical Social Sciences*, **36**, 185–212.

Duggan, J. (1997). "Virtual Bayesian Implementation," *Econometrica*, **65**, 1175–99.

Ehlers, L., H. Peters, and T. Storcken (1999). "Threshold Strategy-proofness," mimeo, University of Maastricht.

Gibbard, A. (1973). "Manipulation of Voting Schemes: A General Result," *Econometrica*, **41**, 587–602.

Gibbard, A. (1977). "Manipulation of Schemes that Mix Voting with Chance," *Econometrica*, **45**, 665–81.

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

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

Harrison, G. and T. McDaniel (2001). "Voting Games and Computational Complexity," mimeo, Univ. of South Carolina.

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.

Hylland, A. (1980). "Strategy Proofness of Voting Procedures with Lotteries as Outcomes and Infinite Sets of Strategies," Harvard University mimeo.

Kalai, E. (2001). "Macro Games," mimeo, Northwestern University.

Kelly, J. (1993). "Almost all Social Choice Rules are Highly Manipulable, but a Few Aren't," *Social Choice and Welfare*, **10**, 161–75.

Moore, J. (1996). "Implementation, Contracts, and Renegotiation in Environments with Complete Information," in *Advances in Economic Theory*, ed. by J.-J. Laffont. Cambridge: Cambridge University Press, 182–282.

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

Roberts, D. and A. Postlewaite (1976). "The Incentives for Price-taking Behavior in Large Exchange Economies," *Econometrica*, **44**, 113–27.

Roth, A. (1982). "Incentive Compatibility in a Market with Indivisible Goods," *Economics Letters*, **9**, 127–32.

Rustichini, A., M. Satterthwaite, and S. Williams (1994). "Convergence to Efficiency in a Simple Market with Incomplete Information," *Econometrica*, **62**, 1041–63.

Saari, D. (1995). *Basic Geometry of Voting*, Springer, Berlin.

Satterthwaite, M. (1975). "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, **10**, 187–217.

Satterthwaite, M. (2001). "Strategy-Proofness and Markets," *Social Choice and Welfare*, **18**, 37–58.

Schummer, J. (1997). "Strategy-proofness versus Efficiency on Restricted Domains of Exchange Economies," *Social Choice and Welfare*, **14**, 47–56.

Schummer, J. (1999). "Strategy-proofness versus Efficiency for Small Domains of Preferences over Public Goods," *Economic Theory*, **13**, 709–22.

SCHUMMER, J., AND W. THOMSON (1997). "Two Derivations of the Uniform Rule and an Application to Bankruptcy," *Economics Letters*, **55**, 333-7.

SCHUMMER, J., AND R. VOHRA (2001). "Strategy-proof Location on a Network," forthcoming in *Journal of Economic Theory.*

SHAPLEY, L., AND H. SCARF (1974). "On Cores and Indivisibility," *Journal of Mathematical Economics*, **1**, 23–37.

SMITH, D. (1999). "Manipulability Measures of Common Social Choice Functions," *Social Choice and Welfare*, **16**, 639–61.

SPRUMONT, Y. (1991). "The Division Problem with Single-Peaked Preferences: A Characterization of the Uniform Allocation Rule," *Econometrica*, **59**, 509–19.

SWINKELS, J. (2001). "Efficiency of Large Private Value Auctions," *Econometrica*, **69**, 37–68.

THOMSON, W. (1994a). "Consistent Solutions to the Problem of Fair Division when Preferences are Single-Peaked," *Journal of Economic Theory*, **63**, 219–45.

THOMSON, W. (1994b). "Resource-monotonic Solutions to the Problem of Fair Division when Preferences are Single-Peaked," *Social Choice and Welfare*, **11**, 205–23.

THOMSON, W. (1995). "Population-monotonic Solutions to the Problem of Fair Division when Preferences are Single-Peaked," *Economic Theory*, **5**, 229–46.

THOMSON, W. (1998). "Strategy-proof Allocation Rules," mimeo, University of Rochester.

ZHOU, L. (1991). "Inefficiency of Strategy Proof Allocation Mechanisms in Pure Exchange Economies," *Social Choice and Welfare*, **8**, 247–54.

Appendix for Referees–Not to be published

9.1 Derivation of utility of agent 2 under φ_L

First we derive the probability of X at each of the two lotteries $y(\succsim_1)$ and $z(\succsim_1)$. Denote those probabilities y_x and z_x. Recall that we are only examining the case such that $X \succ_1 Y \succ_1 Z$.

When $y(\succsim_1) = y'(\succsim_1)$, agent 1's utility from $y(\succsim_1)$ is equal to his utility from ℓ^X. Therefore,

$$y_x \cdot 1 + (1 - y_x) \cdot \lambda_1 = (1 - 2\gamma) \cdot 1 + \gamma\lambda_1 + \gamma \cdot 0$$

which solves as

$$y_x = \frac{1 - 2\gamma + (\gamma - 1)\lambda_1}{1 - \lambda_1}$$

If $y() \neq y'()$, then by definition $y_x = \gamma$.

Similarly, when $z() = z'()$, $z_x = 1 - 2\gamma + \gamma\lambda_1$. In fact in the case we examine here, $z = z'$ always.

Still restricting attention to the case $X \succ_1 Y \succ_2 Z$, the rule φ_L, as a function of agent 2's preferences and of λ_1, is described by the following table.

preference profile	Efficient set	outcome under φ
$X \succsim_2 Y \succsim_2 Z$	$\{x\}$	X
$X \succsim_2 Z \succsim_2 Y$	$\{x\}$	X
$Y \succsim_2 X \succsim_2 Z$	$[x, y]$	$y(\succsim_1)$
$Y \succsim_2 Z \succsim_2 X$	$[x, y]$	$y(\succsim_1)$
$Z \succsim_2 X \succsim_2 Y$	$[x, z]$	$z(\succsim_1)$
$Z \succsim_2 Y \succsim_2 X$ and $(\lambda_1 + \lambda_2 < 1)$	$[x, z]$	$z(\succsim_1)$
$Z \succsim_2 Y \succsim_2 X$ and $(\lambda_1 + \lambda_2 > 1)$	$[x, y] \cup [x, z]$	$y(\succsim_1)$

Given the above derivation of $y()$ and $z()$, it is now straightforward to calculate agent 2's utility as a function of his preferences, λ_2, and λ_1, for the case $X \succsim_1 Y \succsim_1 Z$. The other cases are symmetric.