Theoretical Economics 2 (2007), 203–229 1555-7561/20070203

Secure implementation

TATSUYOSHI SAIJO
Research Institute for Sustainability Science and Institute of Social and Economic Research, Osaka University

TOMAS SJÖSTRÖM
Department of Economics, Rutgers University

TAKEHIKO YAMATO
Graduate School of Decision Science and Technology, Tokyo Institute of Technology

Strategy-proofness, requiring that truth-telling be a dominant strategy, is a standard concept in social choice theory. However, this concept has serious drawbacks. In particular, many strategy-proof mechanisms have multiple Nash equilibria, some of which produce the wrong outcome. A possible solution to this problem is to require double implementation in Nash equilibrium and in dominant strategies, i.e., secure implementation. We characterize securely implementable social choice functions and investigate the connections with dominant strategy implementation and robust implementation. We show that in standard quasi-linear environments with divisible private or public goods, there exist surplus-maximizing (non-dictatorial) social choice functions that can be securely implemented.

KEYWORDS. Nash implementation, robust implementation, secure implementation, strategy-proofness.

JEL CLASSIFICATION. C92, D71, D78, H41.

1. INTRODUCTION

Strategy-proofness, requiring that truth-telling be a dominant strategy, is a standard concept in social choice theory. Although it seems natural that an agent will tell the truth if it is a dominant strategy to do so, there are some problems. First, announcing

Tatsuyoshi Saijo: saijo@iser.osaka-u.ac.jp
Tomas Sjöström: tsjostrom@economics.rutgers.edu
Takehiko Yamato: yamato@valdes.titech.ac.jp

We would like to thank two anonymous referees and especially the co-editor, Jeffrey Ely, for many suggestions that have improved the paper. We are also grateful for helpful comments provided by Salvador Barberà, Dirk Bergemann, Ken Binmore, Hervé Moulin, Shinji Ohseto, Thomas Palfrey, Jun Wako, and Hirofumi Yamamura. Research was partially supported by the Japanese Ministry of the Environment through the project numbered H-062, the Grant in Aid for Scientific Research 15310023 of the Ministry of Education, Science and Culture in Japan, the Japan Securities Scholarship Foundation, and the Abe Fellowship.

Copyright © 2007 Tatsuyoshi Saijo, Tomas Sjöström, and Takehiko Yamato. Licensed under the Creative Commons Attribution-NonCommercial License 2.5. Available at http://econtheory.org.

one's true preference may not be a *unique* dominant strategy, and using the wrong dominant strategy may lead to the wrong outcome. Second, many strategy-proof mechanisms have multiple Nash equilibria, some of which produce the wrong outcome. Third, experimental evidence shows that some strategy-proof mechanisms do not work well; that is, very few subjects reveal their true valuations. For example, see Attiyeh et al. (2000) and Kawagoe and Mori (2001) for pivotal mechanism experiments, and Kagel et al. (1987) and Kagel and Levin (1993) for second-price auction experiments with independent private values.

The first problem can be solved by requiring "full" implementation in dominant strategies. That is, all dominant strategy equilibria should yield a socially optimal outcome. This may require the use of indirect mechanisms. However, Repullo (1985) shows that if a social choice function f is dominant strategy implemented by some indirect mechanism, but f is not dominant strategy implemented by its associated direct mechanism, then the indirect mechanism does not Nash implement f. This leads to the second problem: mechanisms for dominant strategy implementation may have "bad" Nash equilibria. For this reason, Repullo (1985) suggested that the concept of dominant strategy implementation should be replaced by Nash or Bayesian–Nash implementation. We agree that the existence of "bad" (Bayesian) Nash equilibria is problematic. However, in the absence of a dominant strategy, a player's best response depends on the other players' choices, which may be hard to predict. This strategic uncertainty may lead to a failure to coordinate on a (Bayesian) Nash equilibrium. Moreover, a problematic aspect of Bayesian–Nash implementation is that it typically requires the mechanism designer to know the common prior of the players.

It seems clear that the standard concepts—dominant strategy implementation and (Bayesian) Nash implementation—cannot provide a robust foundation for practical implementation. However, if a mechanism simultaneously implements a social choice function in dominant strategies *and* in Nash equilibria, then we get dual advantages. First, with dominant strategies, strategic uncertainty is not important. Second, the mechanism "robustly" implements the social choice function in Bayesian–Nash equilibria, with no need to assume the mechanism designer knows the players' prior beliefs.

A social choice function is *securely implementable* if there exists a game form that simultaneously implements it in dominant strategy equilibria and in Nash equilibria. Thus, all Nash equilibria should yield a socially optimal outcome. We characterize securely implementable social choice functions: a social choice function is securely implementable if and only if it satisfies strategy-proofness and a new property called the *rectangular property*. We show that many quasi-linear economic environments with continuous private or public goods admit securely implementable non-dictatorial social choice functions that maximize social surplus. However, in a standard single-peaked voting model without side-payments, any securely implementable social choice rule must be either dictatorial or Pareto inefficient. This negative result holds even for multi-valued social choice correspondences. In a quasi-linear environment with a *discrete* social decision, such as whether or not to implement an indivisible public project, some interesting non-dictatorial social choice correspondences can be securely implemented, but none of them maximizes the social surplus.

Our hope is that secure implementation may lead to some progress on the third problem mentioned above, the rather negative experimental evidence. We consider secure implementation to be a benchmark: if secure mechanisms do not work well in experiments, then there is very little hope that anything will work. But if a secure mechanism works well in experiments while implementation using less demanding equilibrium concepts fails, then we may be able to pinpoint the reason for the failure by comparing the failed experiment with the benchmark of secure implementation. The question of whether secure mechanisms work well in experiments is investigated in a companion paper (Cason et al. 2006).

The remainder of the paper is organized as follows. We give notation and definitions in Section 2. We characterize secure implementability in Section 3. In Section 4 we discuss the relationship between non-bossiness, dominant strategy implementation, and secure implementation. In Section 5, we consider "robust" Bayesian–Nash implementation. In Section 6, we show the possibility of secure implementation in economies with quasi-linear preferences and divisible public and private goods. Sections 7 and 8 discuss the difficulty of secure implementation with discrete social decisions, and in the absence of side-payments. Sections 2–8 focus on pure strategies. Section 9 extends the analysis to mixed and correlated strategies. Concluding remarks are in Section 10.

2. Notation and definitions

Let A be an arbitrary set of alternatives and let $I = \{1, 2, \dots, n\}$ be the set of agents, with generic element i. We assume that $n \geq 2$. Each agent i has a preference relation defined over A which admits a numerical representation $u_i : A \to \mathbb{R}$. For each i, let U_i be the class of possible utility functions for agent i. Let $u = (u_1, \dots, u_n) \in U \equiv \times_{i \in I} U_i$.

A *social choice function* (SCF) is a function $f : U \to A$ that associates with every $u \in U$ a unique alternative $f(u)$ in A.

A *mechanism* (or *game form*) is a function $g : S \to A$ that assigns to every $s \in S$ a unique element of A, where $S = \times_{i \in I} S_i$ and S_i is the *strategy space of agent* i. The mechanism g is called the *direct revelation mechanism associated with the SCF* f if $S_i = U_i$ for all $i \in I$ and $g(u) = f(u)$ for all $u \in U$. We sometimes abuse terminology by not distinguishing between the SCF f and the direct revelation mechanism associated with f. The list $s \in S$ is written as (s_i, s_{-i}), where $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n) \in S_{-i} \equiv \times_{j \neq i} S_j$. Given $s \in S$ and $s'_i \in S_i$, (s'_i, s_{-i}) is the list $(s_1, \dots, s_{i-1}, s'_i, s_{i+1}, \dots, s_n)$ obtained by replacing the i-th component of s by s'_i. Let $g(S_i, s_{-i})$ be the *attainable set of agent* i *at* s_{-i}, i.e., the set of outcomes that agent i can induce when the other agents select s_{-i}.

For $i \in I$, $u_i \in U_i$, and $a \in A$, let $L(a, u_i) \equiv \{b \in A \mid u_i(a) \geq u_i(b)\}$ be the *weak lower contour set for agent* i *with* u_i *at* a. Given a mechanism $g : S \to A$, the strategy profile $s \in S$ is a *Nash equilibrium of* g *at* $u \in U$ if for all $i \in I$, $g(S_i, s_{-i}) \subseteq L(g(s), u_i)$. Let $N^g(u)$ be the set of Nash equilibria of g at u. The mechanism g *implements the SCF* f *in Nash equilibria* if for each $u \in U$, (i) there exists $s \in N^g(u)$ such that $g(s) = f(u)$ and (ii) for any $s \in N^g(u)$, $g(s) = f(u)$. The SCF f is *Nash implementable* if there exists a mechanism that implements f in Nash equilibria.

Let a mechanism $g : S \to A$ be given. The strategy $s_i \in S_i$ is a *dominant strategy for agent $i \in I$ of g at $u_i \in U_i$* if for all $\hat{s}_{-i} \in S_{-i}$, $g(S_i, \hat{s}_{-i}) \subseteq L(g(s_i, \hat{s}_{-i}), u_i)$. Let $DS_i^g(u_i)$ be the set of dominant strategies for i of g at u_i. The strategy profile $s \in S$ is a *dominant strategy equilibrium of g at $u \in U$* if for all $i \in I$, $s_i \in DS_i^g(u_i)$. Let $DS^g(u)$ be the set of dominant strategy equilibria of g at u. The mechanism g *implements the SCF f in dominant strategy equilibria* if for each $u \in U$, (i) there exists $s \in DS^g(u)$ such that $g(s) = f(u)$ and (ii) for any $s \in DS^g(u)$, $g(s) = f(u)$. The SCF f is *dominant strategy implementable* if there exists a mechanism that implements f in dominant strategy equilibria.

The SCF f is *strategy-proof* if for all $i \in I$, all u_i, $\tilde{u}_i \in U_i$, and all $\tilde{u}_{-i} \in U_{-i}$, $u_i(f(u_i, \tilde{u}_{-i})) \geq u_i(f(\tilde{u}_i, \tilde{u}_{-i}))$. The following result, due to Gibbard (1973), is well-known.

PROPOSITION 1 (Revelation Principle for Dominant Strategy Implementation). *If the SCF f is dominant strategy implementable, then f is strategy-proof.*

Strategy-proofness of the SCF f implies the existence of a mechanism, i.e., the direct revelation mechanism, such that there exists at least one dominant-strategy equilibrium whose outcome is f-optimal at each possible preference profile. On the other hand, dominant strategy implementability requires that *all* dominant strategy equilibria produce the f-optimal outcome. Therefore, the converse of Proposition 1 is not true: some strategy-proof SCF's cannot be dominant strategy implemented (e.g., Dasgupta et al. 1979).

3. SECURE IMPLEMENTATION: A CHARACTERIZATION AND A REVELATION PRINCIPLE

We introduce the following new concept of implementation.

DEFINITION 1. The mechanism g *securely implements the SCF f* if for each $u \in U$, (i) there exists $s \in DS^g(u)$ such that $g(s) = f(u)$ and (ii) for any $s \in N^g(u)$, $g(s) = f(u)$. The SCF f is *securely implementable* if there exists a mechanism that securely implements f.

Secure implementation requires that for every possible preference profile, (i) there exists at least one dominant strategy equilibrium whose outcome is f-optimal and (ii) all Nash equilibria produce the f-optimal outcome.[1]

Next we characterize the class of securely implementable SCF's. We use two conditions. The first condition is strategy-proofness. As Proposition 1 indicates, strategy-proofness is necessary for dominant strategy implementation, and so it is also necessary for secure implementation. However, an additional condition is also necessary for secure implementation. To see why intuitively, suppose that the direct revelation mechanism $g = f$ securely implements the SCF f. See Figure 1 in which $n = 2$ and (u_1, u_2) is

[1]Secure implementation is identical with *double* implementation in dominant strategy equilibria and Nash equilibria. It was Maskin (1979) who first introduced the concept of double implementation. (See also Yamato 1993.) Note that secure implementation can be regarded as multiple (more than double) implementation in dominant strategy equilibria, Nash equilibria, and all refinements of Nash equilibria whose sets are larger than the set of dominant strategy equilibria.

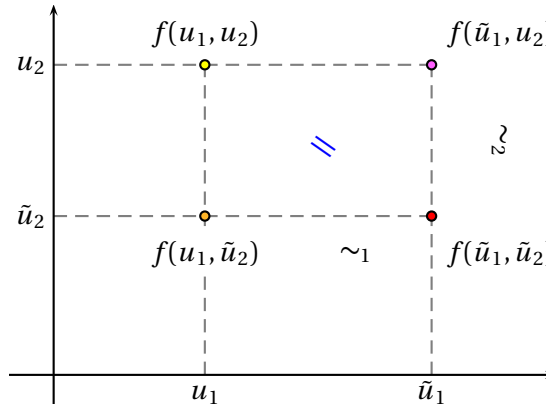

FIGURE 1. The rectangular property.

the true preference profile. Suppose

$$u_1(f(u_1, \tilde{u}_2)) = u_1(f(\tilde{u}_1, \tilde{u}_2)), \tag{1}$$

that is, agent 1 is indifferent between reporting the true preference u_1 and another preference $\tilde{u}_1$ when agent 2's report is $\tilde{u}_2$. Since reporting u_1 is a dominant strategy by strategy-proofness, it follows from (1) that

$$u_1(f(\tilde{u}_1, \tilde{u}_2)) = u_1(f(u_1, \tilde{u}_2)) \geq u_1(f(u_1', \tilde{u}_2)) \text{ for all } u_1' \in U_1.$$

That is, reporting $\tilde{u}_1$ is one of agent 1's best responses at u_1 when agent 2 reports $\tilde{u}_2$.

Next suppose that

$$u_2(f(\tilde{u}_1, u_2)) = u_2(f(\tilde{u}_1, \tilde{u}_2)). \tag{2}$$

By using an argument similar to the one above, it is easy to see that $u_2(f(\tilde{u}_1, \tilde{u}_2)) \geq u_2(f(\tilde{u}_1, u_2'))$ for all $u_2' \in U_1$, that is, reporting $\tilde{u}_2$ is one of agent 2's best responses when agent 1 reports $\tilde{u}_1$. Therefore, $f(\tilde{u}_1, \tilde{u}_2)$ is a Nash equilibrium outcome. Moreover, $f(u_1, u_2)$ is a dominant strategy outcome, and by secure implementability, the dominant strategy outcome coincides with the Nash equilibrium outcome. Accordingly we conclude that $f(u_1, u_2) = f(\tilde{u}_1, \tilde{u}_2)$ if (1) and (2) hold.

A formal definition of this condition, called the *rectangular property*, is given as follows.

DEFINITION 2. The SCF f satisfies the *rectangular property* if for all u, $\tilde{u} \in U$, if $u_i(f(\tilde{u}_i, \tilde{u}_{-i})) = u_i(f(u_i, \tilde{u}_{-i}))$ for all $i \in I$, then $f(\tilde{u}) = f(u)$.

A formal proof of the claim that the rectangular property is necessary for secure implementation is given as follows.

LEMMA 1. *If the SCF f is securely implementable, then f satisfies the rectangular property.*

PROOF. Let $g : S \to A$ be a mechanism that securely implements f. Take any $u, \tilde{u} \in U$. Suppose that

$$u_i(f(\tilde{u}_i, \tilde{u}_{-i})) = u_i(f(u_i, \tilde{u}_{-i})) \text{ for all } i \in I. \tag{3}$$

Choose a dominant strategy profile at $\tilde{u}$, $s(\tilde{u}) = (s_1(\tilde{u}_1), \dots, s_n(\tilde{u}_n)) \in DS^g(\tilde{u})$. By dominant implementability,

$$g(s_1(\tilde{u}_1), \dots, s_n(\tilde{u}_n)) = f(\tilde{u}). \tag{4}$$

Let $i \in I$ be given. Choose a dominant strategy for i at u_i, $s_i(u_i) \in DS_i^g(u_i)$. Then $(s_i(u_i), s_{-i}(\tilde{u}_{-i})) \in DS^g(u_i, \tilde{u}_{-i})$, where $s_{-i}(\tilde{u}_{-i}) = (s_j(\tilde{u}_j))_{j \neq i}$. By dominant implementability,

$$g(s_i(u_i), s_{-i}(\tilde{u}_{-i})) = f(u_i, \tilde{u}_{-i}). \tag{5}$$

By (3), (4), and (5),

$$u_i(g(s_i(u_i), s_{-i}(\tilde{u}_{-i}))) = u_i(g(s_1(\tilde{u}_1), \dots, s_n(\tilde{u}_n))). \tag{6}$$

Further, since $s_i(u_i) \in DS_i^g(u_i)$,

$$g(S_i, s_{-i}(\tilde{u}_{-i})) \subseteq L(g(s_i(u_i), s_{-i}(\tilde{u}_{-i})), u_i). \tag{7}$$

By (6) and (7), $g(S_i, s_{-i}(\tilde{u}_{-i})) \subseteq L(g(s_i(\tilde{u}_i), s_{-i}(\tilde{u}_{-i})), u_i)$. Since this holds for any $i \in I$, $(s_1(\tilde{u}_1), \dots, s_n(\tilde{u}_n)) \in N^g(u)$. By Nash implementability and (4), we have $f(u) = g(s_1(\tilde{u}_1), \dots, s_n(\tilde{u}_n)) = f(\tilde{u})$. □

Next we show that strategy-proofness and the rectangular property are not only necessary, but also sufficient for secure implementability.

LEMMA 2. *If the SCF f satisfies strategy-proofness and the rectangular property, then the direct revelation mechanism associated with f securely implements f.*

PROOF. Consider the direct revelation mechanism $g = f$. Let $u \in U$ be given. By strategy-proofness, $u \in DS^f(u)$, that is, the truthful strategy profile is a dominant strategy equilibrium of the direct revelation mechanism. Next we prove that for any $\tilde{u} \in N^f(u)$, $f(\tilde{u}) = f(u)$, that is, any Nash equilibrium of the direct revelation mechanism produces the f-optimal outcome. Since $\tilde{u} \in N^f(u)$,

$$u_i(f(\tilde{u}_i, \tilde{u}_{-i})) \geq u_i(f(u_i, \tilde{u}_{-i})) \text{ for all } i \in I. \tag{8}$$

Further, since $u_i \in DS_i^f(u_i)$ by strategy-proofness,

$$u_i(f(u_i, \tilde{u}_{-i})) \geq u_i(f(\tilde{u}_i, \tilde{u}_{-i})) \text{ for all } i \in I. \tag{9}$$

By (8) and (9), $u_i(f(\tilde{u}_i, \tilde{u}_{-i})) = u_i(f(u_i, \tilde{u}_{-i}))$ for all $i \in I$. By the rectangular property, $f(\tilde{u}) = f(u)$. □

By Proposition 1 and Lemmas 1 and 2, we have the following characterization of securely implementable SCF's.

THEOREM 1. *An SCF is securely implementable if and only if it satisfies strategy-proofness and the rectangular property.*

In the early literature on implementation, it was pointed out that even if an SCF f is implementable in dominant strategies, it may not be implemented by its associated direct revelation mechanism: it may be necessary to use more complicated "indirect" mechanisms (Dasgupta et al. 1979, Repullo 1985). However, the same is not true for a securely implementable mechanism. Suppose the SCF f is securely implemented by some mechanism. Then by Proposition 1 and Lemma 1, f satisfies strategy-proofness and the rectangular property. Hence by Lemma 2, f is securely implemented by its associated direct revelation mechanism. Thus, we have a *revelation principle for secure implementation.*

THEOREM 2. *An SCF is securely implementable if and only if it is securely implemented by its associated direct revelation mechanism.*

The implication of this revelation principle is that we can limit our attention to the set of direct mechanisms. Direct mechanisms are somewhat natural and easy to explain to experimental subjects, which may add to their appeal.

4. NON-BOSSINESS, DOMINANT STRATEGY IMPLEMENTATION, AND SECURE IMPLEMENTATION

To further study the set of securely implementable social choice functions, we need the idea of *non-bossiness*. Intuitively, non-bossiness implies that no one can change the outcome without changing her own utility. Satterthwaite and Sonnenschein (1981) first introduced a definition of non-bossiness for economic environments.[2] For general environments, consider the following definition.

DEFINITION 3. The SCF f satisfies *non-bossiness* if for all $u, u' \in U$ and all $i \in I$, if $f(u_i, u_{-i}) \neq f(u'_i, u_{-i})$, then $u_i(f(u_i, u_{-i})) \neq u_i(f(u'_i, u_{-i}))$.

PROPOSITION 2. *If an SCF satisfies the rectangular property, then it satisfies non-bossiness.*

PROOF. Suppose the SCF f satisfies the rectangular property, and $u_j(f(u'_j, u_{-j})) = u_j(f(u_j, u_{-j}))$ for some j. Let u'' be such that $u'' = (u'_j, u_{-j})$. We need to show $f(u'') = f(u)$. Now $u_j(f(u'')) = u_j(f(u'_j, u_{-j})) = u_j(f(u_j, u_{-j})) = u_j(f(u_j, u''_{-j}))$, and $(u''_i, u''_{-i}) = (u_i, u''_{-i})$ for all $i \neq j$. So we have $u_i(f(u''_i, u''_{-i})) = u_i(f(u_i, u''_{-i}))$ for all $i \in I$. By the rectangular property, $f(u'') = f(u)$. □

[2]Our definition of non-bossiness is slightly stronger than Satterthwaite–Sonnenschein's original condition when applied to economic environments. Satterthwaite and Sonnenschein's original definition is that the SCF f satisfies non-bossiness if for all $u, u' \in U$ and all $i \in I$, if $f(u_i, u_{-i}) \neq f(u'_i, u_{-i})$, then $f_i(u_i, u_{-i}) \neq f_i(u'_i, u_{-i})$, where $f_i(u)$ denotes the consumption bundle agent i receives at the allocation $f(u) = (f_i(u))_{i \in I}$ recommended by the SCF f for the preference profile u. Mizukami and Wakayama (2007) discuss the importance of non-bossiness for dominant strategy implementation in exchange economies.

The rectangular property is stronger than non-bossiness. However, the rectangular property is equivalent to non-bossiness plus the following weak version of the rectangular property.[3]

DEFINITION 4. The SCF f satisfies the *outcome-rectangular-property (ORP)* if for all u, $u' \in U$, if $f(u_i, u'_{-i}) = f(u')$ for all $i \in I$, then $f(u) = f(u')$.

PROPOSITION 3. *An SCF satisfies the rectangular property if and only if it satisfies non-bossiness and ORP.*

PROOF. By Proposition 2, the rectangular property implies non-bossiness. It is also clear that the rectangular property implies ORP. Next suppose the SCF f satisfies non-bossiness and ORP, and $u_i(f(u'_i, u_{-i})) = u_i(f(u_i, u_{-i}))$ for all i. Then it follows from non-bossiness that $f(u'_i, u_{-i}) = f(u_i, u_{-i})$ for all i. By ORP, $f(u') = f(u)$. Therefore, f satisfies the rectangular property. □

By Theorem 1 and Proposition 3, we have the following corollary.

COROLLARY 1. *An SCF is securely implementable if and only if it satisfies strategy-proofness, non-bossiness, and ORP.*

Thus any securely implementable SCF must be non-bossy. On the other hand, there are non-bossy and strategy-proof SCFs that violate ORP, hence cannot be securely implemented (an example is provided in Section 6). However, it turns out that the following condition of *weak* non-bossiness is enough to guarantee that a strategy-proof SCF can be *dominant strategy implemented*.

DEFINITION 5. The SCF f satisfies *weak non-bossiness* if for all $u, u' \in U$ and all $i \in I$, if $f(u_i, u_{-i}) \neq f(u'_i, u_{-i})$, then there is some u''_{-i} such that $u_i(f(u_i, u''_{-i})) \neq u_i(f(u'_i, u''_{-i}))$.

THEOREM 3. *An SCF is dominant strategy implemented by its associated direct revelation mechanism if and only if it satisfies strategy-proofness and weak non-bossiness.*

PROOF. Suppose the SCF f satisfies strategy-proofness and weak non-bossiness. Consider the associated direct revelation mechanism. Suppose agent i's true preference is u_i. By strategy proofness, it is dominant to announce the truth u_i. Suppose announcing a different preference u'_i is another dominant strategy. If $f(u_i, u_{-i}) \neq f(u'_i, u_{-i})$ for some u_{-i}, then by weak non-bossiness there is u''_{-i} such that $u_i(f(u_i, u''_{-i})) > u_i(f(u'_i, u''_{-i}))$. Therefore, announcing u'_i is in fact dominated by announcing u_i, which is a contradiction. Hence, $f(u_i, u_{-i}) = f(u'_i, u_{-i})$ for all u_{-i} after all, so agent i's lie (i.e. to say u'_i) cannot ever affect the outcome. Hence, f is dominant strategy implemented.

Suppose the SCF f is dominant strategy implemented by its associated direct revelation mechanism. By Proposition 1, f is strategy-proof. It remains to show f satisfies weak non-bossiness. Take any $u, u' \in U$ and $i \in I$. Suppose $f(u_i, u_{-i}) \neq f(u'_i, u_{-i})$. Then announcing u'_i is dominated by announcing u_i when agent i's true preference is u_i, so that there is u''_{-i} such that $u_i(f(u_i, u''_{-i})) > u_i(f(u'_i, u''_{-i}))$. □

[3]We thank an anonymous referee for suggesting this condition.

Non-bossiness is a stronger condition than weak non-bossiness, so secure implementation is more difficult to achieve than dominant strategy implementation. For example, the Vickrey auction discussed in Section 7 satisfies weak non-bossiness, but violates non-bossiness. (Notice that in general, weak non-bossiness does not imply that each player has a unique dominant strategy in the revelation mechanism.)

5. Robust Bayesian implementation

The standard theory of Bayesian mechanism design makes the strong assumption that the agents and the mechanism designer share a common prior over the possible states of the world. The mechanism can depend directly on this prior, i.e., it can be "parametric." For example, in Myerson's (1981) optimal auction, the mechanism designer sets a reserve price that depends on the prior distribution. In this section, we drop this strong assumption and consider the possibility of "non-parametric" implementation.[4]

In a novel approach, Bergemann and Morris (2005a,b) allow each agent to have a set of possible belief-types (as well as payoff types), and define equilibrium with respect to these extended type-spaces. We assume also that, from the point of view of the mechanism designer, each agent has a set of possible beliefs. But we define equilibrium with respect to the agents' actual beliefs.[5] We are interested mainly in the standard case of a common prior, in which case the equilibrium is defined with respect to this common prior. Our definitions are, however, more general, and allow beliefs to differ across agents.

The set of possible utility functions for agent i is a measurable space U_i. Let $U = \times_{i\in I} U_i$. A preference profile $u \in U$ is referred to as a "state of the world." Consider a mechanism $g : S \to A$. A strategy for player i is a measurable function $\sigma_i : U_i \to S_i$, with the following interpretation: when player i's true preference relation is u_i, he plays $\sigma_i(u_i)$. A strategy profile is a measurable function $\sigma : U \to S$, where $\sigma(u) = (\sigma_1(u_1), \sigma_2(u_2), \dots, \sigma_n(u_n))$. Similarly, define $\sigma_{-i}(u_{-i})$ in the obvious way. Then a dominant strategy equilibrium is a strategy profile σ such that for each i and each $u \in U$,

$$u_i(g(\sigma_i(u_i), \tilde{\sigma}_{-i}(u_{-i})) \geq u_i(g(\tilde{\sigma}_i(u_i), \tilde{\sigma}_{-i}(u_{-i})))$$

for any $\tilde{\sigma}_i$ and $\tilde{\sigma}_{-i}$. Notice that the notion of dominant strategy does not depend on beliefs.

Let Δ be the set of probability measures over U, and let $\Delta^n = \Delta \times \cdots \times \Delta$. Agent i's prior belief over U is denoted q_i $(i = 1, \dots, n)$. A *support* for q_i is any set C_i such that $q_i(C_i) = 1$. Let $D \subseteq \Delta^n$ be the set of prior belief profiles $(q_1, q_2, \dots, q_n)$ that the agents could possibly have. (We allow the possibility that $q_i \neq q_j$ for any agents $i, j \in I$.) The mechanism designer does not know the agents' true priors; she knows only that they belong to the set D. Several special classes of priors are of interest. Let D^{CI} be the class of *complete information priors*:

$$D^{CI} = \{(q_1, q_2, \dots, q_n) \in \Delta^n : \exists u \text{ s.t. } q_1(u) = q_2(u) = \cdots = q_n(u) = 1\}.$$

[4]The arguments in this section have benefited from suggestions from the co-editor.

[5]In addition, unlike Bergemann and Morris (2005a,b) we assume "private values."

Let D^{COM} be the class of *common priors*:

$$D^{COM} = \{(q_1, q_2, \ldots, q_n) \in \Delta^n : q_1 = q_2 = \cdots = q_n\}.$$

Given priors $(q_i)_{i\in I}$, the strategy profile σ is a Bayesian–Nash equilibrium under $(q_i)_{i\in I}$ if for all i,

$$\int_U u_i(g(\sigma(u)))\, dq_i(u) \geq \int_U u_i(g(\delta_i(u_i), \sigma_{-i}(u_{-i})))\, dq_i(u)$$

for any alternative strategy δ_i.

Our notion of robust implementation requires the outcome to be optimal in each state in the support of all agents' priors. Importantly, since the mechanism designer does not know the agents' true priors, the mechanism must be "non-parametric," i.e., the *same* mechanism must achieve implementation *for all priors in the set* D.[6]

DEFINITION 6. The mechanism g *securely and robustly implements the SCF* f *on the domain* D if for any collection of priors $(q_1, q_2, \ldots, q_n) \in D$, (i) there exists a dominant strategy equilibrium σ such that $g(\sigma(u)) = f(u)$ for all $u \in U$ and (ii) for any Bayesian–Nash equilibrium σ under $(q_1, q_2, \ldots, q_n)$, there is a support C_i for q_i $(i = 1, 2, \ldots, n)$ such that $g(\sigma(u)) = f(u)$ for all $u \in \cap_i C_i$.

Consider, for example, the special case $D = D^{CI}$. In this case, the agents have complete information. Since the mechanism designer knows that the priors $(q_1, q_2, \ldots, q_n)$ belong to D^{CI}, she knows that $q_1(u) = q_2(u) = \cdots = q_n(u) = 1$ for *some* $u \in U$. But she does not know *which* u. According to her, any profile of priors in D^{CI} is a priori possible. Now if $u \in U$ is such that $q_1(u) = q_2(u) = \cdots = q_n(u) = 1$, then any support C_i must include this u, so part (ii) of Definition 6 requires that $g(\sigma(u)) = f(u)$ for any Bayesian–Nash equilibrium σ. That is, if all agents agree that the true state is u, then the outcome must be socially optimal for this u. Furthermore, σ is a Bayesian–Nash equilibrium under $(q_1, q_2, \ldots, q_n)$ if and only if $\sigma(u)$ is a Nash equilibrium at u. This implies that Definition 6 generalizes Definition 1. Specifically, the mechanism g *securely and robustly implements the SCF* f *on the domain* D^{CI} (according to Definition 6) if and only if g *securely implements* f (according to Definition 1).

Definition 6 is of course more general than Definition 1, since it covers cases where $D \neq D^{CI}$. For example, we may consider a polar opposite to the case of complete information. Suppose U is finite and D contains only *full support* priors: $(q_1, q_2, \ldots, q_n) \in D$ implies $q_i(u) > 0$ for all $u \in U$ and all $i \in I$. In this case, $C_i = U$ for all $i \in I$, so part (ii) of Definition 6 requires $g(\sigma(u)) = f(u)$ for all $u \in U$. That is, if the agents' priors do not

[6]Suppose the mechanism designer has her own prior q_o on U, which may differ from the agents' priors, but any event the mechanism designer considers possible is also considered possible by the agents. Formally, q_o is absolutely continuous with respect to all of the agents' priors. If the mechanism designer wants the outcome to be optimal with probability one according to q_o, it suffices if the outcome is optimal for states in $\cap_i C_i$, where C_i is a support for q_i. (If U is finite then $C_i \subseteq U$ is a support for q_i if and only if it includes all u such that $q_i(u) > 0$. In this case, the mechanism designer's prior q_o is absolutely continuous with respect to $(q_i)_{i\in I}$ if for any $u \in U$, $q_0(u) > 0$ implies $q_i(u) > 0$ for all $i \in I$.)

rule out any state as being impossible, then the outcome must be socially optimal for all $u \in U$.

A final example is the standard case of common priors, $D = D^{COM}$. In this case, part (ii) of Definition 6 requires that the outcome be socially optimal with probability one according to the common prior, whatever it may be.

Our main theorem in this section is the following result.

THEOREM 4. *If the SCF f is strategy-proof and satisfies the rectangular property, then it is securely and robustly implemented on any D by its associated direct revelation mechanism.*

PROOF. Suppose f satisfies strategy-proofness and the rectangular property, and consider the associated direct revelation mechanism. Fix $(q_1, q_2, \dots, q_n) \in D$. Strategy-proofness implies that the truthful strategy profile, $\sigma^*(u) = u$ for any $u \in U$, is a Bayesian–Nash equilibrium. Next consider *any* Bayesian–Nash equilibrium σ. By definition, for any $i \in I$,

$$\int_U u_i(f(\sigma(u)))\, dq_i(u) \geq \int_U u_i(f(u_i, \sigma_{-i}(u_{-i})))\, dq_i(u).$$

That is, playing according to σ_i is at least good as always telling the truth. On the other hand, by strategy-proofness,

$$u_i(f(u_i, \sigma_{-i}(u_{-i}))) \geq u_i(f(\sigma(u)))$$

for all $u \in U$. These inequalities imply that there is a support C_i for q_i such that

$$u_i(f(u_i, \sigma_{-i}(u_{-i}))) = u_i(f(\sigma(u)))$$

for all $u \in C_i$. Then for any $u \in \cap_{i \in I} C_i$, the above equality holds for any agent i. By the rectangular property, $f(\sigma(u)) = f(u)$ for all $u \in \cap_i C_i$. □

Thus, strategy-proofness (i.e. dominant-strategy incentive compatibility) and the rectangular property are sufficient for secure and robust implementation. For a partial converse, recall that secure and robust implementation on D^{CI} is equivalent to secure implementation. This observation and Theorem 2 regarding necessary conditions for secure implementation immediately imply the following result.

THEOREM 5. *Suppose $D^{CI} \subseteq D$. If the SCF f can be securely and robustly implemented on D, then f is strategy-proof and satisfies the rectangular property.*

Equilibrium behavior requires agents to coordinate on a strategy profile. Since such coordination may be difficult in practice, we impose part (i) of Definition 6. However, for theoretical reasons it may be of interest to weaken Definition 6. We say that the mechanism g *robustly implements the SCF f in Bayesian–Nash equilibria* if condition (i) of secure and robust implementation is replaced by the following condition: (i′) there

exist a Bayesian–Nash equilibrium σ and a support C_i for q_i $(i = 1,2,\dots,n)$ such that $g(\sigma(u)) = f(u)$ for all $u \in \cap_i C_i$.

Robust implementation is easier to achieve than secure and robust implementation. Consider, for example, the special case $D = D^{CI}$. If $(q_1, q_2, \dots, q_n) \in D^{CI}$, then $q_1(u) = q_2(u) = \cdots = q_n(u) = 1$ for some $u \in U$, and C_i is a support for q_i $(i = 1,2,\dots,n)$ if and only if $u \in \cap_i C_i$. Moreover, σ is a Bayesian–Nash equilibrium under $(q_1, q_2, \dots, q_n)$ if and only if $\sigma(u)$ is a Nash equilibrium at u. Therefore, the mechanism g *robustly implements the SCF f on the domain D^{CI}* if and only if g *Nash implements f*. Robust implementation for the set of all complete information priors is, naturally, logically equivalent to Nash implementation. Given some U, let f be any Nash implementable SCF (for example, one that satisfies Maskin-monotonicity and no-veto-power). Then, f is robustly implementable on D^{CI}. Yet, if f is not strategy-proof then f is not securely and robustly implementable on D^{CI}. In general, therefore, secure and robust implementation is a more demanding notion than robust implementation. Of course, the bigger is the set D, the less likely it is that an SCF that is not strategy-proof can be Bayesian–Nash implemented by a "non-parametric" mechanism.

With very general type spaces, Bergemann and Morris (2005a) show that robust implementation essentially boils down to iterated elimination of strictly dominated strategies. More positive results have been obtained in less general settings. Choi and Kim (1999) consider the case $D = D^{COM}$ and find that a well-known SCF that is not strategy-proof can be implemented in *undominated* Bayesian–Nash equilibria using a non-parametric mechanism. This result cannot be directly translated into our setting, because their notion of SCF is different from ours. According to their SCF, the socially optimal outcome depends on the agents' prior beliefs as well as their payoff functions.

A precise characterization of social choice rules that are robustly Bayesian–Nash implementable, under various assumptions on D, is beyond the scope of this paper. However, we note that the mechanism used by Choi and Kim (1999) is quite complex. A key aspect of their mechanism is that *the agents announce their common prior*. In any *undominated* Bayesian–Nash equilibrium, the agents report their priors *truthfully*, which allows the mechanism designer to extract information about the agents' true prior beliefs. The extent to which such mechanisms can achieve Bayesian–Nash (as opposed to *undominated* Bayesian–Nash) implementation is unclear, since ruling out Bayesian–Nash equilibria where the agents all announce the "wrong" prior may be quite difficult. In any case, from the point of view of practical implementation, it is interesting to see what can be achieved using simple "revelation mechanisms."[7] Thus, suppose $g : U \to U$, and replace condition (i) of Definition 6 by: (i″) there exists a truthful Bayesian–Nash equilibrium σ (i.e., $\sigma(u) = u$) and a support C_i for q_i $(i = 1,2,\dots,n)$ such that $g(\sigma(u)) = f(u)$ for all $u \in \cap_i C_i$. The resulting concept of implementation is called *robust and truthful* implementation.

Dasgupta et al. (1979, Theorem 5.1) show that if truth-telling is a Bayesian–Nash equilibrium for all possible common priors, then it must be a dominant strategy. Their

[7] A broad interpretation of revelation mechanisms requires the agents to report all they know, including their own prior beliefs, as in Choi and Kim (1999). Here we use a more narrow definition, where the agents reveal only their "payoff types."

argument carries over to our setting. Consequently, robust and truthful implementation on D^{CI} implies secure implementation (because telling the truth must be a dominant strategy). This observation, combined with Theorem 2, immediately implies the following result.

THEOREM 6. *Suppose $D^{CI} \subseteq D$. If f can be robustly and truthfully implemented in Bayesian–Nash equilibria, then f is strategy-proof and satisfies the rectangular property (so f can be securely and robustly implemented on D).*

Theorem 6 implies that weakening Definition 6 by replacing (i) by (i″) does not really impact the possibility of implementation. As remarked earlier, condition (i′) may yield more permissive results (depending on the exact nature of D), but a full characterization is left for future work.

6. QUASI-LINEAR ECONOMIC ENVIRONMENTS

Let the set of alternatives be

$$A = \{(y, t_1, \dots, t_n) \mid y \in Y, t_i \in \mathbb{R} \ \forall i\},$$

where $y \in Y$ is a social decision and t_i is a transfer to agent i of a private good called "money." The set of possible social decisions Y is a convex subset of $\mathbb{R}^k$, for some k. (In the next section, we consider the case where Y is a discrete set.) The cost of taking decision y (in terms of "money") is given by a differentiable and convex function $c(y)$. Each agent $i \in I$ has quasi-linear preferences:

$$u_i(y, t_1, \dots, t_n, \theta_i) = v_i(y, \theta_i) + t_i.$$

Here v_i is a valuation function that is differentiable and concave in y, and θ_i is a real number representing agent i's "type."[8] For each i, the function v_i is given once and for all and only the type varies, so the preferences of the agents are represented by the profile of types, $\theta = (\theta_1, \theta_2, \dots, \theta_n)$. The set of possible types for agent i is Θ_i. Let $\Theta \equiv \times_{i \in I} \Theta_i$. An SCF $f : \Theta \to A$ recommends, for each profile θ, a social decision $y^f(\theta)$ and a set of transfers. Let $t_i^f(\theta)$ denote the recommended transfer to agent i. Thus, $f(\theta) = (y^f(\theta), t_1^f(\theta), t_2^f(\theta), \dots, t_n^f(\theta))$. The *social surplus* is defined as

$$\sum_{i \in I} v_i(y, \theta_i) - c(y). \tag{10}$$

To avoid some technical issues, in this section we assume that for all $\theta \in \Theta$, a unique y maximizes the social surplus (10). (This happens, for example, if each v_i is strictly concave in y.) A direct revelation mechanism $f : \Theta \to A$ is a *Groves–Clarke mechanism* if

[8]Notice that throughout this section, the functions u_i and v_i are held fixed, and the agent's type is identified with θ_i. The mechanism designer knows the functions u_i and v_i but not the agent's true type. Moreover, each agent i's valuation function v_i depends only on his own type θ_i, and not the other agents' types $(\theta_j)_{j \neq i}$, since we consider private values environments.

for all $\theta \in \Theta$, $y^f(\theta)$ maximizes the social surplus, and the transfer function is given, for all $i \in I$, by

$$t_i^f(\theta) = \sum_{j \neq i} v_j(y^f(\theta), \theta_j) - c(y^f(\theta)) + \varphi_i(\theta_{-i}). \tag{11}$$

Here φ_i is an arbitrary function that does not depend on θ_i. It is well-known that Groves–Clarke mechanisms are strategy-proof (Clarke 1971, Groves 1973). In many cases, for example if each v_i is differentiable in θ_i and each Θ_i is a convex set, any strategy-proof SCF that satisfies (10) must in fact also satisfy (11) (Holmström 1979).

If the social surplus maximizing decision always occurs in the interior of Y (denoted $\operatorname{int} Y$) then the rectangular property is equivalent to non-bossiness. Both properties reduce to the following: no agent should be able to change the profile of transfers without changing the social decision. This is shown in the following lemma.

LEMMA 3. *Suppose for all $\theta \in \Theta$, $y^f(\theta) \in \operatorname{int} Y$ maximizes the social surplus. For any Groves–Clarke mechanism $f : \Theta \to A$, the following three conditions are equivalent: (i) f is non-bossy; (ii) for all $\theta, \theta' \in \Theta$ and $i \in I$, $f(\theta) = f(\theta_i', \theta_{-i})$ whenever $y^f(\theta) = y^f(\theta_i', \theta_{-i})$; (iii) f satisfies the rectangular property.*

PROOF. (i) implies (ii). Strategy proofness implies that if $y^f(\theta) = y^f(\theta_i', \theta_{-i})$, then $t_i^f(\theta) = t_i^f(\theta_i', \theta_{-i})$. Non-bossiness then implies $f(\theta) = f(\theta_i', \theta_{-i})$.

(ii) implies (iii). Suppose (ii) holds. Take any $\theta, \theta' \in \Theta$, and let $y^* = y^f(\theta')$. Suppose $u_i(f(\theta_i, \theta_{-i}'), \theta_i) = u_i(f(\theta'), \theta_i)$, i.e., $v_i(y^f(\theta_i, \theta_{-i}'), \theta_i) + t_i^f(\theta, \theta_{-i}') = v_i(y^f(\theta'), \theta_i) + t_i^f(\theta')$ for all i. Then, by (11), $v_i(y^f(\theta_i, \theta_{-i}'), \theta_i) + \sum_{j \neq i} v_j(y^f(\theta_i, \theta_{-i}'), \theta_j') - c(y^f(\theta_i, \theta_{-i}')) = v_i(y^f(\theta'), \theta_i) + \sum_{j \neq i} v_j(y^f(\theta'), \theta_{-j}') - c(y^f(\theta'))$. That is, at state (θ_i, θ_{-i}'), $y^f(\theta_i, \theta_{-i}')$ and $y^f(\theta')$ generate exactly the same social surplus. Since a unique y maximizes the social surplus (10) and $y^f(\theta_i, \theta_{-i}')$ is the maximizer at (θ_i, θ_{-i}'), it follows from this equality that $y^f(\theta_i, \theta_{-i}') = y^f(\theta') = y^*$. By property (ii), $f(\theta') = f(\theta_i, \theta_{-i}')$. Since y^* is interior and v is differentiable and concave in y, y^* can be found by solving the first-order condition for maximizing (10). Since $y^f(\theta_i, \theta_{-i}') = y^f(\theta') = y^*$, we have $\partial v_i(y^*, \theta_i)/\partial y = \partial v_i(y^*, \theta_i')/\partial y$ for all i.

We know that $f(\theta_i, \theta_{-i}') = f(\theta')$ for all i. For $i = 1$, this yields $f(\theta_1, \theta_2', \ldots, \theta_n') = f(\theta')$. The first-order condition for maximizing (10) must be satisfied at y^* for the profile $(\theta_1, \theta_2', \ldots, \theta_n')$. Since $\partial v_2(y^*, \theta_2)/\partial y = \partial v_2(y^*, \theta_2')/\partial y$, the first-order condition is also satisfied at y^* for the profile $(\theta_1, \theta_2, \theta_3', \ldots, \theta_n')$. Thus

$$y^f(\theta_1, \theta_2, \theta_3', \ldots, \theta_n') = y^f(\theta_1, \theta_2', \theta_3', \ldots, \theta_n') = y^*.$$

Property (ii) now implies

$$f(\theta_1, \theta_2, \theta_3', \ldots, \theta_n') = f(\theta_1, \theta_2', \theta_3', \ldots, \theta_n') = f(\theta').$$

By sequentially replacing each θ_i' by θ_i in this manner, we find that $f(\theta) = f(\theta')$. Therefore, the rectangular property holds.

(iii) implies (i). This follows from Proposition 2. □

Example 1 shows that standard assumptions often guarantee non-bossiness.

EXAMPLE 1 (Production of a divisible public good). The public good is one-dimensional, $Y = \mathbb{R}_+$. Two leading cases have been studied in the literature. *Case 1*: $v_i(y, \theta_i) = \theta_i b(y)$, where b is a strictly concave function. To guarantee that the surplus maximizing y is strictly positive, suppose $b'(0) > 0$ and $c'(0) = 0$. *Case 2*: Let $g(x)$ be a function that is strictly concave, reaching a maximum at $x = 0$, and suppose $v_i(y, \theta_i) = g(y - \theta_i)$. There is no cost of producing the public good, $c(y) = 0$. This is the case of single-peaked preferences, where θ_i is agent i's "peak," i.e., his most preferred level of the public good. As long as all θ_i are strictly positive, the surplus maximizing level of the public good is strictly positive.

In both case 1 and case 2, if $y^f(\theta) = y^f(\theta_i', \theta_{-i})$ then $\theta_i' = \theta_i$, so obviously $f(\theta) = f(\theta_i', \theta_{-i})$. From Lemma 3 it follows that all Groves–Clarke mechanisms are non-bossy and securely implement the social surplus maximizing decision. ◊

Example 2 shows that corner solutions do not necessarily mean that secure implementation is impossible.

EXAMPLE 2 (Allocation of a divisible private good in fixed supply). One unit of a divisible private good called "cake" is to be shared by the agents. (In addition, transfers of "money" are possible.) The social decision is denoted $y = (y_1, y_2, \dots, y_n)$, where y_i is agent i's share of the cake. Feasibility requires $y \geq 0$ and $\sum_i y_i = 1$. Valuation functions are of the form $v_i(y, \theta_i) = \theta_i b(y_i)$, where b is a strictly increasing and strictly concave function satisfying $b(0) = 0$. Suppose $\Theta_i = [\theta^{\min}, \theta^{\max}]$, where

$$\theta^{\min} b'(0) > \theta^{\max} b'(1). \tag{12}$$

Inequality (12) guarantees that the social surplus is not maximized by giving all of the cake to one agent. However, with three or more agents, it may be optimal to give no cake to some agent, so Lemma 3 does not apply. The social surplus $\sum_i \theta_i b(y_i)$ is to be maximized subject to $y \geq 0$ and $\sum_i y_i = 1$. Let $\lambda > 0$ denote the Lagrange multiplier for the resource constraint. The maximum is found by solving the first-order condition,

$$\theta_i b'(y_i) \leq \lambda, \quad y_i \geq 0, \quad y_i(\lambda - \theta_i b'(y_i)) = 0 \quad \text{for all } i. \tag{13}$$

Suppose the function φ_i in (11) is a constant, so the transfer of money to agent i is

$$t_i^f(\theta) = \sum_{j \neq i} \theta_j b(y_j^f(\theta)) + \text{constant}. \tag{14}$$

We claim that in this case the Groves–Clarke mechanism satisfies the rectangular property. Indeed, suppose $u_i(f(\theta)) = u_i(f(\theta_i', \theta_{-i}))$ for all i. This implies that for all i, either $\theta_i' = \theta_i$ or agent i gets no cake, $y_i^f(\theta_i', \theta_{-i}) = y_i^f(\theta) = 0$. Therefore, the first-order condition (13) still holds when θ is replaced by θ', without changing λ or y, so $y^f(\theta') = y^f(\theta)$. Moreover, (14) implies $t^f(\theta') = t^f(\theta)$, so $f(\theta') = f(\theta)$ (recall that $b(0) = 0$). Thus, the rectangular property holds, and secure implementation is achieved. ◊

Example 2 illustrates the difference between implementation in *strictly* dominant strategies and secure implementation. In Example 2, telling the truth is not a strictly dominant strategy, because an agent who gets no cake may still get no cake—and the same transfer of money—after a small change in his type. However, this does not prevent secure implementation, as long as the change in his type does not change anyone else's transfer. This is why φ_i must be constant. (If φ_i is not constant then it can happen that $t^f(\theta') \neq t^f(\theta)$ even though $y^f(\theta') = y^f(\theta)$.)

If (12) does not hold, then the Groves–Clarke mechanism with constant φ_i is still non-bossy. However, ORP is violated. Since one agent may consume all of the cake when (12) is violated, $u_i(f(\theta)) = u_i(f(\theta_i', \theta_{-i}))$ implies either $\theta_i' = \theta_i$ or $y_i^f(\theta_i', \theta_{-i}) = y_i^f(\theta) = 0$ or $y_i^f(\theta_i', \theta_{-i}) = y_i^f(\theta) = 1$. But this no longer ensures that the first-order condition (13) holds when θ is replaced by θ'. Therefore, $f(\theta') \neq f(\theta)$ is possible. Intuitively, there can be bad Nash equilibria where one agent exaggerates his valuation of cake and receives all of it, while all the other agents report very low valuations and receive no cake. Notice that this example shows that, in general, non-bossiness and strategy-proofness together do not imply the rectangular property.

EXAMPLE 3 (Serial cost sharing). The social decision is $y = (y_1, y_2, \ldots, y_n)$, where y_i is agent i's consumption of divisible "cake." But in contrast to the assumption of Example 2, cake can now be produced (using money as input). The cost function is $c(y) = C\left(\sum_i y_i\right)$, where C is strictly increasing, differentiable, and convex. Each valuation function v_i is strictly increasing and strictly concave in y_i (but does not depend on y_j for $j \neq i$). Moulin and Shenker (1992) define *serial cost sharing* and show that this SCF is strategy-proof and can be Nash implemented by an indirect mechanism. In general, the two properties of Nash implementability and strategy-proofness together do not imply the rectangular property (which requires double implementation by the same mechanism). However, serial cost sharing does satisfy the rectangular property. Suppose $u_i(f(\theta^*)) = u_i(f(\theta_i, \theta_{-i}^*))$ for all $i \in I$. The definition of serial cost sharing implies $f(\theta^*) = f(\theta_i, \theta_{-i}^*)$ for all $i \in I$. This implies that if $y_i^f(\theta^*) > 0$ then $\partial v_i(y^f(\theta^*), \theta_i)/\partial y_i = \partial v_i(y^f(\theta^*), \theta_i^*)/\partial y_i$. If $y_i^f(\theta^*) = 0$, then $\partial v_i(y^f(\theta^*), \theta_i)/\partial y_i \leq C'\left(\sum_j y_j^f(\theta^*)\right)$ and $\partial v_i(y^f(\theta^*), \theta_i^*)/\partial y_i \leq C'\left(\sum_j y_j^f(\theta^*)\right)$. In either case, $f(\theta^*) = f(\theta)$, so serial cost-sharing is securely implementable. Notice that in this example, there is no need for any assumptions that rule out corner solutions. ◊

7. DISCRETE SOCIAL DECISIONS

The previous section shows that surplus-maximizing social choice functions can be securely implemented in many quasi-linear environments with divisible public or private goods. In this section, we show that secure implementation is more difficult if the set of social decisions is discrete. Consider a quasi-linear environment as in Section 6, but assume that Y is a finite set. For convenience, $Y = \{0, 1\}$ and $c(0) = c(1) = 0$. (The arguments can be adapted to any discrete Y.) We normalize so that $v_i(0, \theta_i) = 0$ for all θ_i. Thus, agent i's preferences are characterized by $v_i(1, \theta_i)$, the value to him of social

decision $y = 1$. Without loss of generality we may suppose $v_i(1, \theta_i) = \theta_i$ for all θ_i. We assume θ_i can be any real number.

Notice that if by chance $\sum_{i \in I} \theta_i = 0$, then both $y = 0$ and $y = 1$ are surplus maximizing. In this situation, it may be unreasonable to assume that the social choice rule is single-valued. Thus, we allow f to be a multi-valued *social choice correspondence* (SCC). The definition of secure implementation when f is an SCC is the same as the one in Definition 1. (Thus, we require "full" implementation in dominant strategy equilibria and Nash equilibria.) Notice that for implementation in *strictly* dominant strategies, the issue of multi-valuedness would be moot because a strictly dominant strategy must be unique. However, in this paper we consider domination in the weak sense, and a given type of player may have several (weakly) dominant strategies. Moreover, even if each player has a unique dominant strategy, there may be multiple Nash equilibria (some of which are in dominated strategies). Secure implementation does not require a unique Nash equilibrium, but it does require that all Nash equilibrium outcomes be socially optimal (see Example 4 below).

We again use the notation $f(\theta) = (y^f(\theta), t_1^f(\theta), t_2^f(\theta), \dots, t_n^f(\theta))$, but now $y^f(\theta)$ and $t_i^f(\theta)$ are the *sets* of optimal decisions and transfers, respectively. The SCC f is *surplus maximizing* if $\sum_{i \in I} \theta_i < 0$ implies $y^f(\theta) = \{0\}$, and $\sum_{i \in I} \theta_i > 0$ implies $y^f(\theta) = \{1\}$. No restriction is imposed if $\sum_{i \in I} \theta_i = 0$. For a mechanism $g : S \to A$, let $g(s) = (y^g(s), t^g(s))$ denote the outcome, where $y^g(s)$ is the chosen public project and $t^g(s)$ the profile of transfers.

To see that some interesting social choice correspondences can be securely implemented, even with a discrete set of public decisions, consider the following "veto rule."

EXAMPLE 4 (A veto rule). Consider the following SCC. There are no transfers. The public decision $y = 0$ is always socially optimal. The public decision $y = 1$ is socially optimal if and only if $\theta_i \geq 0$ for all i. Intuitively, $y = 0$ is a "status quo" outcome, which is always socially acceptable, but the social project $y = 1$ is acceptable to society if and only if *all* agents prefer it to the status quo. (With this interpretation, the SCC is the "individually rational" correspondence.) This SCC is securely implemented by the following mechanism. Each agent says 0 or 1. If all say 1, $y = 1$ is implemented. If at least one agent says 0, then $y = 0$ is implemented. Notice that the dominant strategy is to say 0 if $\theta_i < 0$ and 1 if $\theta_i > 0$. Both strategies are dominant if $\theta_i = 0$. There are no "bad" Nash equilibria, because each agent can "veto" the outcome $y = 1$. The "veto rule" is non-dictatorial: for each agent i, there is a profile θ such that agent i strictly prefers $y = 1$, but the unique socially optimal decision is $y = 0$. However, it does not maximize the surplus, because $y = 0$ is socially acceptable even if $\theta_i > 0$ for all i. ◊

We now show that in fact surplus maximization cannot be achieved in this environment. Notice that this negative result holds, even though we do not require budget balance (i.e., $\sum_{i \in I} t_i \neq 0$ is allowed).

THEOREM 7. *Consider the quasi-linear environment with $Y = \{0, 1\}$. No SCC is both securely implementable and surplus-maximizing.*

PROOF. Suppose f is surplus-maximizing. In order to obtain a contradiction, suppose it is securely implemented by a mechanism g.

Fix a type profile θ and choose $s_j \in DS_j^g(\theta_j)$ for each j. Surplus maximization implies that for any i, $y^g(s) = 0$ if θ_i satisfies

$$\theta_i < -\sum_{j \neq i} \theta_j. \tag{15}$$

If θ_i satisfies

$$\theta_i > -\sum_{j \neq i} \theta_j \tag{16}$$

then $y^g(s) = 1$. Moreover, if (15) holds, then any $s_i \in DS_i^g(\theta_i)$ must give agent i the same transfer, say $t_i^g(s) = t_i^0(s_{-i})$. (Otherwise, the strategy that gives the lowest transfer and the same public decision $y^g(s) = 0$ is not a dominant strategy.) Similarly, if (16) holds, then any $s_i \in DS_i^g(\theta_i)$ must give agent i the same transfer, say $t_i^g(s) = t_i^1(s_{-i})$.

Suppose θ is such that $\sum_{i \in I} \theta_i > 0$. Define a new profile θ' as follows. For $i \in \{1,2\}$, define $\theta_i' = -\sum_{j \neq i} \theta_j - \varepsilon < \theta_i$, where $\varepsilon > 0$. Let $\theta_i' = \theta_i$ for all $i > 2$. For each i, choose $s_i' \in DS_i^g(\theta_i')$. Clearly, $\sum_{i \in I} \theta_i' < 0$. Moreover, for $i \in \{1,2\}$, $\theta_i + \sum_{j \neq i} \theta_j' < 0$. For all i, we have chosen $s_i \in DS_i^g(\theta_i)$ and $s_i' \in DS_i^g(\theta_i')$. By surplus maximization, $y^g(s') = 0$ and $y^g(s_i, s_{-i}') = 0$ for $i \in \{1,2\}$. We now claim that, for $i \in \{1,2\}$, if agent i's true type is θ_i then $s_i' \in DS_i^g(\theta_i')$ is a best response against s_{-i}'. Indeed, choosing s_i' would result in payoff $t_i^0(s_{-i}')$, because the social decision would be $y^g(s') = 0$. But this is also what is obtained by choosing $s_i \in DS_i^g(\theta_i)$, because $y^g(s_i, s_{-i}') = 0$. Therefore, s_i' is indeed a best response against s_{-i}' for $i \in \{1,2\}$ when his true type is θ_i. For all $i > 2$, $\theta_i' = \theta_i$ and $s_i' \in DS_i^g(\theta_i') = DS_i^g(\theta_i)$. Therefore, $s' \in N^g(\theta)$. But $y^g(s') = 0$ even though $\sum_{i \in I} \theta_i > 0$, which contradicts the definition of surplus maximization. □

Notice that the proof of Theorem 7 in effect replicates the proof that the rectangular property is necessary for secure implementation, and then shows that the rectangular property is violated.

To further illustrate the impossibility of combining secure implementation with surplus maximization in the discrete environment, we consider a well-known example.

EXAMPLE 5 (Auctioning an indivisible object). Suppose the social decision is to allocate a private *indivisible* object among two agents. Agent i's true value of the object is $\theta_i \geq 0$ if she receives it, and 0 otherwise ($i = 1,2$). Consider the second-price auction (Vickrey 1961). Suppose $\theta_1 > \theta_2 > 0$. In order to maximize the surplus, agent 1 should win the object. Figure 2 shows that the set of Nash equilibria is quite large. The lower-right part of the set of Nash equilibria is the "good set" in the sense that agent 1 receives the object. However, the upper-left part of the set of Nash equilibria is "bad" in the sense that agent 2 receives the object, so the social surplus is not maximized. ◊

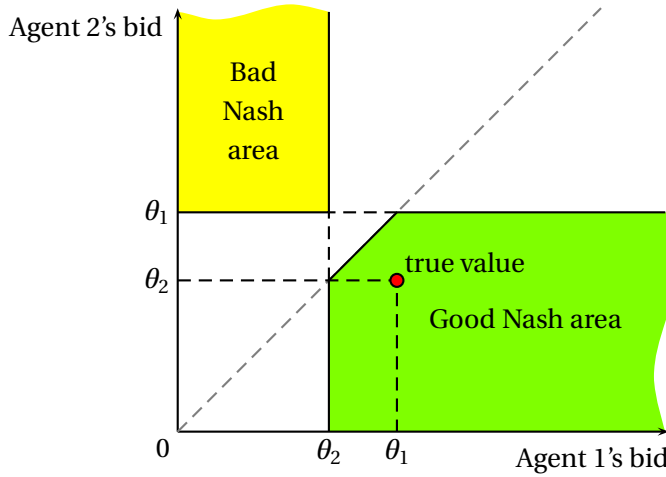

FIGURE 2. Equilibria of the second-price auction.

8. SINGLE-PEAKED VOTING

Section 6 shows the possibility of secure implementation when the social decision is concerned with continuous variables, such as divisible public or private goods. However, the mechanisms rely on the existence of "money" for side-payments. We now show that if there are no side-payments, the results are negative, even if the social decision is a continuous variable.

Consider a *single-peaked voting environment*. The set of alternatives is $A = [0,1]$, and the set of possible preference relations consists of all those that are continuous and single-peaked on A. Let $p(u_i)$ denote the "peak" of u_i, i.e., the top ranked alternative in A, which is assumed to be unique. Single-peakedness implies that u_i is strictly increasing before $p(u_i)$ and strictly decreasing afterwards.[9] Let Range(f) denote the range of f. By Lemma 1 in Barberà and Jackson (1994), Range(f) is closed. Let $a = \min\{x \mid x \in \text{Range}(f)\}$ and $b = \max\{x \mid x \in \text{Range}(f)\}$ denote the smallest and largest elements in Range(f), respectively. Notice that f is constant if and only if $a = b$.

In the single-peaked voting environment one can find dominant strategy implementable social choice functions with good properties, the leading example being the median voter rule (see Barberà and Jackson 1994). Assuming $n \geq 3$ is odd, the median voter rule has many good properties: it is strategy-proof, non-dictatorial, Pareto efficient, and non-bossy. However, it does not satisfy ORP. To see this, suppose u' is such that each u'_i has the same peak, say $p(u'_i) = a$ for all $i \in I$. Then, $f(u') = a$. Now suppose for each i, u_i is such that $p(u) = b \neq a$. Then, $f(u_i, u'_{-i}) = f(u'_i, u'_{-i}) = a$ for all i, because starting at u', no single agent can change the median-voter outcome. But

[9]An impossibility result for secure implementation on domain U a fortiori implies the same impossibility result on any larger domain $U' \supseteq U$. Therefore, all impossibility results in this section hold also for a *weakly* single-peaked voting environment, in which u_i is weakly increasing before $p(u_i)$ and weakly decreasing afterwards.

$f(u) = b \neq f(u')$, so ORP is violated. Therefore, by Corollary 1, the median voter rule cannot be securely implemented. The same argument applies verbatim to other well-known social choice rules, such as the one that always picks the smallest of all peaks. More generally, we show that if a Pareto efficient social choice rule can be securely implemented, then it must be dictatorial. This is true even if we allow the social choice rule to be multi-valued. Before proving these negative results for secure implementation, we prove two lemmas.

LEMMA 4. *Let f be a securely implementable non-constant SCF in the single-peaked voting environment. There is an agent i and an alternative $y \in \mathrm{Range}(f)$, $y > a$, such that $f(u) = y$ whenever $p(u_i) \geq y \geq p(u_j)$ for all $j \neq i$.*

PROOF. Let u' be any profile such that $p(u'_i) = a$ for all i, and let u'' be any profile such that $p(u''_i) = b$ for all i. Strategy-proofness implies $f(u') = a$ and $f(u'') = b$, and $b \neq a$ since f is not constant. If $f(u''_i, u'_{-i}) = a$ for all i, then ORP implies that $f(u'') = a$, but this contradicts $f(u'') = b$. Thus, there is an agent, say agent $i = 1$, such that $f(u''_1, u'_{-1}) > a$. Define $y \equiv f(u''_1, u'_{-1})$.

Now let u_1 be any utility function such that $p(u_1) \geq y$. Consider $f(u_1, u'_{-1})$. If $f(u_1, u'_{-1}) > y$, then $u''_1(f(u_1, u'_{-1})) > u''_1(f(u''_1, u'_{-1}))$, and if $f(u_1, u'_{-1}) < y$, then $u_1(f(u''_1, u'_{-1})) > u_1(f(u_1, u'_{-1}))$. Since in either case we have a contradiction of strategy-proofness, we conclude that $f(u_1, u'_{-1}) = y$.

Now, for each $j \geq 2$, let u_j be any utility function such that $p(u_j) \leq y$. Consider $f(u_1, u_j, u'_{-1,j})$. If $f(u_1, u_j, u'_{-1,j}) > y$, then $u_j(f(u_1, u'_j, u'_{-1,j})) > u_j(f(u_1, u_j, u'_{-1,j}))$, and if $f(u_1, u_j, u'_{-1,j}) < y$, then $u'_j(f(u_1, u_j, u'_{-1,j})) > u'_j(f(u_1, u'_j, u'_{-1,j}))$. Since in either case we have a contradiction of strategy-proofness, we conclude that $f(u_1, u_j, u'_{-1,j}) = y$ for all $j \geq 2$.

Condition ORP implies that $f(u) = y$. Thus, $f(u) = y$ whenever $p(u_1) \geq y \geq p(u_j)$ for all $j \geq 2$. □

LEMMA 5. *Let f be a securely implementable non-constant SCF in the single-peaked voting environment. There is an agent i such that $f(u) = a$ whenever $p(u_i) = a$, and $f(u) = b$ whenever $p(u_i) = b$.*

PROOF. Without loss of generality, suppose agent $i = 1$ is the agent identified in Lemma 4, and y the alternative identified in the same lemma. Let u' be any profile such that $p(u'_i) = a$ for all i, and let $\tilde{u}$ be any profile such that $p(\tilde{u}_i) = y$ for all i. Then $f(u') = a$ by strategy-proofness, and Lemma 4 implies $f(\tilde{u}) = y$. If $f(u'_i, \tilde{u}_{-i}) = y$ for all i, then ORP implies that $f(u') = y$, but this contradicts $f(u') = a$. Thus, there is an agent i such that $f(u'_i, \tilde{u}_{-i}) \neq y$. Lemma 4 implies that in fact $i = 1$. Strategy-proofness implies $f(u'_1, \tilde{u}_{-1}) < y$. Let $z \equiv f(u'_1, \tilde{u}_{-1}) < y$. We show that $z = a$.

It is impossible that $z < a$ because $a = \min\{x : x \in \mathrm{Range}(f)\}$. Suppose $z > a$. Now let $\hat{u}$ be a profile such that $p(\hat{u}_i) = z$ for all i. Strategy-proofness implies $f(\hat{u}) = z$. Since $z = f(u'_1, \tilde{u}_{-1})$, strategy-proofness implies $f(u'_1, \hat{u}_i, \tilde{u}_{-1,i}) = z$ for all $i > 1$. Then ORP implies $f(u'_1, \hat{u}_{-1}) = z$.

Now consider $f(u_i', \hat{u}_{-i})$ for $i > 1$. Strategy-proofness requires $f(u_i', \hat{u}_{-i}) \leq z$. Notice that this inequality holds regardless of $\hat{u}_1$, as long as $p(\hat{u}_1) = z$. Moreover, $f(u_i', \hat{u}_{-i})$ is in fact the same alternative for any $\hat{u}_1$ such that $p(\hat{u}_1) = z$. (Otherwise, there would exist $\hat{u}_1$ and $\bar{u}_1$ such that $p(\hat{u}_1) = p(\bar{u}_1) = z$ and $f(\hat{u}_1, u_i', \hat{u}_{-1,i}) < f(\bar{u}_1, u_i', \hat{u}_{-1,i}) \leq z$. But then $\hat{u}_1(f(\hat{u}_1, u_i', \hat{u}_{-1,i})) < \hat{u}_1(f(\bar{u}_1, u_i', \hat{u}_{-1,i}))$, contradicting strategy-proofness.) Suppose $w \equiv f(u_i', \hat{u}_{-i}) < z < y$. But now consider $\hat{u}_1$ such that $p(\hat{u}_1) = z$ and $\hat{u}_1(y) > \hat{u}_1(w)$. Lemma 4 implies that if $p(\tilde{u}_1) = y$, then $f(\tilde{u}_1, u_i', \hat{u}_{-1,i}) = y$. But since $\hat{u}_1(y) > \hat{u}_1(w)$ and $w = f(u_i', \hat{u}_{-i})$, strategy-proofness is violated. This contradiction implies $f(u_i', \hat{u}_{-i}) = z$ for all $i > 1$. Since we have already established $f(u_1', \hat{u}_{-1}) = z$, we can apply ORP and conclude that $f(u') = z$. However, $f(u') = a$, a contradiction of our hypothesis that $z > a$. So, we must have $z = a$.

The previous paragraph establishes that $f(u_1', \tilde{u}_{-1}) = a$ whenever $p(u_1') = a$ and $p(\tilde{u}_i) = y$ for all $i > 1$. Now for all $i > 1$, let $\tilde{u}_i$ be such that $p(\tilde{u}_i) = y$ and $\tilde{u}_i(x) > \tilde{u}_i(a)$ for all $x \in \text{Range}(f)$ such that $x > a$. Consider any agent $i > 1$ and any arbitrary u_i. If $f(u_1', u_i, \tilde{u}_{-1,i}) \neq a$, then $\tilde{u}_i(f(u_1', u_i, \tilde{u}_{-1,i})) > \tilde{u}_i(f(u_1', \tilde{u}_i, \tilde{u}_{-1,i}))$, which contradicts strategy-proofness. Hence, $f(u_1', u_i, \tilde{u}_{-1,i}) = a$ for all $i > 1$. ORP implies $f(u_1', u_{-1}) = a$. We conclude that $f(u_1', u_{-1}) = a$ whenever $p(u_1') = a$.

Exactly the same line of reasoning establishes the existence of an agent i such that $f(u_i', u_{-i}) = b$ whenever $p(u_i') = b$. Obviously, this must be $i = 1$, or else we contradict the already established fact that $f(u_1', u_{-1}) = a$ whenever $p(u_1') = a$. □

Now we are ready to prove our first negative result for single-peaked voting. It covers the case of single-valued social choice rules.

THEOREM 8. *Let f be a securely implementable SCF in the single-peaked voting environment. There is a dictator on* Range(f)*, i.e., an agent i such that for all u and all $x \in$ Range(f), $u_i(f(u)) \geq u_i(x)$.*

PROOF. Since the result is trivial if f is constant, suppose f is securely implementable but not constant. Lemma 5 identifies an agent i such that $f(u) = a$ whenever $p(u_i) = a$, and $f(u) = b$ whenever $p(u_i) = b$. Without loss of generality suppose this is true for $i = 1$. Fix any $x \in \text{Range}(f)$. Let u' be such that $p(u_i') = x$ for all i, and let u be an arbitrary profile. The theorem is proved by showing that $f(u_1', u_{-1}) = x$ must necessarily hold.

Strategy-proofness implies $f(u') = x$. Fix any $i > 1$. We show that $f(u_i, u_{-i}') = x$. If $p(u_i) = x$, then $f(u_i, u_{-i}') = x$ by strategy-proofness. Suppose instead that $p(u_i) > x$. Then strategy-proofness implies $f(u_i, u_{-i}') \geq x$. Notice that this inequality holds regardless of u_1', as long as $p(u_1') = x$. Moreover, $f(u_i, u_{-i}')$ is in fact the same alternative for any u_1' such that $p(u_1') = x$. (Otherwise, there would exist u_1' and u_1'' such that $p(u_1') = p(u_1'') = x$ and $f(u_1', u_i, u_{-1,i}') > f(u_1'', u_i, u_{-1,i}') \geq x$. But then $u_1'(f(u_1', u_i, u_{-1,i}')) < u_1'(f(u_1'', u_i, u_{-1,i}'))$, contradicting strategy-proofness.)

Now suppose $z \equiv f(u_i, u_{-i}') > x$. But consider u_1' such that $p(u_1') = x$ and $u_1'(a) > u_1'(z)$. If $\tilde{u}_1$ is such that $p(\tilde{u}_1) = a$, then $f(\tilde{u}_1, u_i, u_{-1,i}') = a$ by Lemma 5. But then $u_1'(f(\tilde{u}_1, u_i, u_{-1,i}')) > u_1'(f(u_1', u_i, u_{-1,i}'))$, contradicting strategy-proofness. This contradiction shows that we must have $f(u_i, u_{-i}') = x$ whenever $p(u_i) > x$. A similar argument

establishes that $f(u_i, u'_{-i}) = x$ whenever $p(u_i) < x$. We conclude that, for all $i > 1$, $f(u_i, u'_{-i}) = x$ for all u_i. ORP implies $f(u'_1, u_{-1}) = x$. □

As in the previous section, there exist non-dictatorial social choice correspondences that can be securely implemented. For example, a "veto rule," similar to Example 4, with some arbitrary alternative designated as status quo, can be securely implemented in the single-peaked voting model. However, this SCC is not Pareto efficient. More generally, in this environment an SCC is either single-valued, in which case it is dictatorial by Theorem 8, or it is Pareto inefficient. This is our second negative result for single-peaked voting.

THEOREM 9. *Let f be a securely implementable SCC in the single-peaked voting environment. Then f is either single-valued or Pareto inefficient.*

PROOF. Suppose f is a securely implementable SCC that is not single-valued. Then there is u such that $f(u)$ contains at least two distinct alternatives. If f is securely implemented by mechanism g, then there are two strategy profiles $s, s' \in DS^g(u)$ such that $g(s) \neq g(s')$. Then, there necessarily exist alternatives a and b, and an agent i, such that $g(s'_1, \dots, s'_{-i}, s_i, s_{i+1}, \dots, s_n) = a$ but $g(s'_1, \dots, s'_{-i}, s'_i, s_{i+1}, \dots, s_n) = b \neq a$. We may choose labeling so that $i = 1$ and $b > a$.

Thus, we have $s \in DS^g(u)$, $g(s) = a$, $(s'_1, s_{-1}) \in DS^g(u)$, and $g(s'_1, s_{-1}) = b > a$. Since $s_1, s'_1 \in DS^g_1(u_1)$, it must be the case that $a < p(u_1) < b$ and $u_1(a) = u_1(b)$.

Let $L = \{j : p(u_j) \leq a\}$ be the set of agents whose peaks, in the profile u, are (weakly) to the left of a. Suppose $2 \in L$, and suppose u^*_2 is such that $a < p(u^*_2) < b$ and $u^*_2(a) > u^*_2(b)$. Let $s^*_2 \in DS^g_2(u^*_2)$.

CLAIM. $g(s_1, s^*_2, s_{-1,2}) = a$.

PROOF. To prove the claim, we consider the various possibilities.

Case 1: $a < g(s_1, s^*_2, s_{-1,2}) < b$. Since $s_1, s'_1 \in DS^g_1(u_1)$, we have $u_1(g(s_1, s^*_2, s_{-1,2})) = u_1(g(s'_1, s^*_2, s_{-1,2}))$. Therefore, $a < g(s'_1, s^*_2, s_{-1,2}) < b$. But $g(s'_1, s_2, s_{-1,2}) = b$ and $2 \in L$, so $u_2(g(s'_1, s^*_2, s_{-1,2})) > u_2(g(s'_1, s_2, s_{-1,2}))$. However, this contradicts $s_2 \in DS^g_2(u_2)$. Therefore, case 1 is impossible.

Case 2: $g(s_1, s^*_2, s_{-1,2}) < a = g(s)$. This case is impossible because $p(u^*_2) > a$ and $s^*_2 \in DS^g_2(u^*_2)$.

Case 3: $g(s_1, s^*_2, s_{-1,2}) \geq b$. But then, $u^*_2(g(s)) > u^*_2(b) \geq u^*_2(g(s_1, s^*_2, s_{-1,2}))$, which contradicts $s^*_2 \in DS^g_2(u^*_2)$.

Since cases 1, 2, and 3 are all impossible, the claim is true.

The claim establishes $g(s_1, s^*_2, s_{-1,2}) = a$. Since $s_1, s'_1 \in DS^g_1(u_1)$, it must be the case that $u_1(g(s'_1, s^*_2, s_{-1,2})) = u_1(a)$. This means that $g(s'_1, s^*_2, s_{-1,2})$ can be either a or b. Suppose $g(s'_1, s^*_2, s_{-1,2}) = a$. But $g(s'_1, s_{-1}) = b$. Since $2 \in L$ we have $u_2(a) > u_2(b)$, which contradicts $s_2 \in DS^g_2(u_2)$. Therefore, we have $g(s'_1, s^*_2, s_{-1,2}) = b$.

To summarize, we have shown that $(s^*_2, s_{-2}) \in DS^g(u^*_2, u_{-2})$, $g(s^*_2, s_{-2}) = a$, $(s'_1, s^*_2, s_{-1,2}) \in DS^g(u^*_2, u_{-2})$, and $g(s'_1, s^*_2, s_{-1,2}) = b > a$. This puts us back in our original position, except that the L set has one fewer member after u_2 is replaced by u^*_2

(because $p(u_2^*) > a$). We can repeat the same argument for each $j \in L$: we let u_j^* be such that $a < p(u_2^*) < b$ and $u_j^*(a) > u_j^*(b)$, and we pick $s_j^* \in DS_j^g(u_j^*)$. After having exhausted all the members of L, we obtain $s_L^* = \{s_j^*\}_{j\in L}$, where $s_j^* \in DS_j^g(u_j^*)$ for each $j \in L$, and $g(s_{-L}, s_L^*) = a$. Since g securely implements f, $a \in f(u_{-L}, u_L^*)$. However, by definition of L, when the utility profile is (u_{-L}, u_L^*), all agents have peaks strictly to the right of a. Therefore, a is not Pareto efficient. □

9. Correlated equilibria

So far we have restricted attention to pure strategy equilibria following a tradition in the implementation literature.[10] However, showing that mixed strategies do not cause any problems is an important step toward making implementation more secure. In this section, we allow for agents to use mixed or more generally correlated strategies. We show that our characterization results on secure implementation for pure strategy equilibria hold even for correlated strategy equilibria. Bergemann and Morris (2005a) obtain a similar result for their definition of robust implementation. They use a purification argument that relies on their general type-spaces. Our method of proof is different.

Consider a mechanism $g : S \to A$. For notational simplicity, assume both U and S are finite sets. A *correlated strategy profile* for the mechanism is denoted $\mu : U \to \Delta(S)$, with the following interpretation: when the true preference profile is u, the probability the agents play $s \in S$ is $\mu(s|u)$. A *correlated equilibrium* is a correlated strategy profile μ such that for each $u \in U$ and each $i \in I$,

$$\sum_{s\in S} u_i(g(s))\mu(s|u) \geq \sum_{s\in S} u_i(g(h(s_i), s_{-i}))\mu(s|u)$$

for any function $h : S_i \to S_i$.

The mechanism g *securely implements the SCF f in correlated equilibria* if (i) for each $u \in U$, there exists a (pure) strategy profile $s \in DS^g(u)$ such that $g(s) = f(u)$ and (ii) for any correlated equilibrium μ, $g(s) = f(u)$ whenever $\mu(s|u) > 0$. We now extend Theorem 1 to cover secure implementation in correlated equilibria.

Theorem 10. *An SCF is securely implementable in correlated equilibria if and only if it satisfies strategy-proofness and the rectangular property.*

Proof. Since every pure-strategy Nash equilibrium is a correlated equilibrium, "only if" follows from Theorem 1. So consider "if." Suppose that f satisfies strategy-proofness and the rectangular property, and consider the associated direct revelation mechanism. The first requirement of secure implementability is satisfied since by strategy-proofness, the truthful pure-strategy profile, $\sigma(u) = u$, is a pure-strategy dominant strategy equilibrium for any utility profile $u \in U$.

Next we prove that the second requirement is also met. Consider any correlated equilibrium μ of the direct mechanism, and suppose $\mu(\tilde{u}|u) > 0$. We want to show

[10] Exceptions considering mixed strategy equilibria include Jackson (1992), Jackson et al. (1994), Sjöström (1994), and Maskin (1999).

$f(\tilde{u}) = f(u)$. Let $i \in I$ be given. Because μ is a correlated equilibrium, playing according to μ is at least good as telling the truth at the state u:

$$\sum_{u' \in U} u_i(f(u'_i, u'_{-i}))\mu(u'|u) \geq \sum_{u' \in U} u_i(f(u_i, u'_{-i}))\mu(u'|u).$$

On the other hand, by strategy-proofness,

$$u_i(f(u_i, u'_{-i})) \geq u_i(f(u'_i, u'_{-i}))$$

for all $u' \in U$. Together these inequalities imply that for all $u' \in U$ with $\mu(u'|u) > 0$,

$$u_i(f(u_i, u'_{-i})) = u_i(f(u'_i, u'_{-i})) \quad \forall i \in I.$$

Since $\mu(\tilde{u}|u) > 0$, $u_i(f(u_i, \tilde{u}_{-i})) = u_i(f(\tilde{u}_i, \tilde{u}_{-i}))$ for all $i \in I$. By the rectangular property, it follows that $f(\tilde{u}) = f(u)$. □

A mixed-strategy equilibrium is a correlated equilibrium with independent randomizations, i.e., $\mu(s|u) = \prod_i \mu_i(s_i|u_i)$. It is clear that Theorem 10 continues to hold if "correlated equilibria" is replaced by "mixed-strategy equilibria." In a similar way, we can extend our results on secure and robust implementation (and robust and truthful implementation) to cover correlated (or mixed) Bayesian–Nash equilibria.

10. Concluding remarks

Some recent experiments suggest that strategy-proof mechanisms may not work well in practice. Motivated by this finding, we have shown that *strategy-proofness plus the rectangular property* is a necessary and sufficient condition for the following concepts of implementation.

(a) Secure implementation in pure-strategy (or mixed, or correlated) Nash equilibria.

(b) Secure and robust implementation in pure-strategy (or mixed, or correlated) Bayesian–Nash equilibria.

(c) Robust and truthful implementation in pure-strategy (or mixed, or correlated) Bayesian–Nash equilibria.

For (b) and (c), necessity requires $D^{CI} \subseteq D$.

In standard quasi-linear environments with divisible public or private goods, Groves–Clarke mechanisms satisfy both strategy-proofness and the rectangular property, and so are "secure." In other environments, the rectangular property turns out to be surprisingly hard to satisfy. In some such environments, positive results might still be obtained for *partial* implementation of social choice *correspondences*: each player must have a dominant strategy, and every (Bayesian) Nash equilibrium outcome must be socially optimal. However, it is not required that every socially optimal outcome be a dominant strategy equilibrium outcome. This interesting problem is left for future work.

The next step is to investigate if "secure" mechanisms work in practice. In Cason et al. (2006), we report experiments on two strategy-proof mechanisms: the pivotal mechanism with two agents and a binary public project that has a continuum of Nash equilibria, and a Groves–Clarke mechanism with two agents and single-peaked preferences that has a unique Nash equilibrium. We find that subjects played dominant strategies significantly more frequently in the secure Groves mechanism than in the non-secure pivotal mechanism. This makes us optimistic about the future of mechanism design. The negative experimental evidence mentioned in the introduction was based on mechanisms that are not secure (such as the second-price auction). In these experiments, there may have been insufficient pressure on the players to adopt their dominant strategies, and deviations may not have been punished by big payoff losses (for a discussion, see Cason et al. 2006). Imposing stricter requirements than simply strategy-proofness may turn out to be the key to successful applications of mechanism design.

References

Attiyeh, Greg, Robert Franciosi, and R. Mark Isaac (2000), "Experiments with the pivot process for providing public goods." *Public Choice*, 102, 95–114. [204]

Barberà, Salvador and Matthew O. Jackson (1994), "A characterization of strategy-proof social choice functions for economies with pure public goods." *Social Choice and Welfare*, 11, 241–252. [221]

Bergemann, Dirk and Stephen Morris (2005a), "Robust implementation: The role of large type spaces." Discussion Paper 1519, Cowles Foundation, Yale University. [211, 214, 225]

Bergemann, Dirk and Stephen Morris (2005b), "Robust mechanism design." *Econometrica*, 73, 1771–1813. [211]

Cason, Timothy N., Tatsuyoshi Saijo, Tomas Sjöström, and Takehiko Yamato (2006), "Secure implementation experiments: Do strategy-proof mechanisms really work?" *Games and Economic Behavior*, 57, 206–235. [205, 227]

Choi, Jaewon and Taesung Kim (1999), "A nonparametric, efficient public good decision mechanism: Undominated Bayesian implementation." *Games and Economic Behavior*, 27, 64–85. [214]

Clarke, Edward H. (1971), "Multipart pricing of public goods." *Public Choice*, 11, 17–33. [216]

Dasgupta, Partha S., Peter J. Hammond, and Eric S. Maskin (1979), "The implementation of social choice rules: Some general results on incentive compatibility." *Review of Economic Studies*, 46, 185–216. [206, 209, 214]

Gibbard, Allan (1973), "Manipulation of voting schemes: A general result." *Econometrica*, 41, 587–601. [206]

Groves, Theodore (1973), "Incentives in teams." *Econometrica*, 41, 617–631. [216]

Holmström, Bengt (1979), "Groves' scheme on restricted domains." *Econometrica*, 47, 1137–1144. [216]

Jackson, Matthew O. (1992), "Implementation in undominated strategies: A look at bounded mechanisms." *Review of Economic Studies*, 59, 757–775. [225]

Jackson, Matthew O., Thomas R. Palfrey, and Sanjay Srivastava (1994), "Undominated Nash implementation in bounded mechanisms." *Games and Economic Behavior*, 6, 474–501. [225]

Kagel, John H., Ronald M. Harstad, and Dan Levin (1987), "Information impact and allocation rules in auctions with affiliated private values: A laboratory study." *Econometrica*, 55, 1275–1304. [204]

Kagel, John H. and Dan Levin (1993), "Independent private value auctions: Bidder behaviour in first-, second- and third-price auctions with varying numbers of bidders." *Economic Journal*, 103, 868–879. [204]

Kawagoe, Toshiji and Toru Mori (2001), "Can the pivotal mechanism induce truth-telling? An experimental study." *Public Choice*, 108, 331–354. [204]

Maskin, Eric (1979), "Incentive schemes immune to group manipulation." Unpublished paper. [206]

Maskin, Eric (1999), "Nash equilibrium and welfare optimality." *Review of Economic Studies*, 66, 23–38. [225]

Mizukami, Hideki and Takuma Wakayama (2007), "Dominant strategy implementation in economic environments." *Games and Economic Behavior*, 60, 307–325. [209]

Moulin, Hervé and Scott Shenker (1992), "Serial cost sharing." *Econometrica*, 60, 1009–1037. [218]

Myerson, Roger B. (1981), "Optimal auction design." *Mathematics of Operation Research*, 6, 58–73. [211]

Repullo, Rafael (1985), "Implementation in dominant strategies under complete and incomplete information." *Review of Economic Studies*, 52, 223–229. [204, 209]

Satterthwaite, Mark A. and Hugo Sonnenschein (1981), "Strategy-proof allocation mechanisms at differentiable points." *Review of Economic Studies*, 48, 587–597. [209]

Sjöström, Tomas (1994), "Implementation in undominated Nash equilibria without integer games." *Games and Economic Behavior*, 6, 502–511. [225]

Vickrey, William (1961), "Counterspeculation, auctions, and competitive sealed tenders." *Journal of Finance*, 16, 8–37. [220]

Yamato, Takehiko (1993), "Double implementation in Nash and undominated Nash equilibria." *Journal of Economic Theory*, 59, 311–323. [206]

Submitted 2006-3-29. Final version accepted 2007-5-13. Available online 2007-5-14.