Essays on Strategy-proofness and Implementation

Sonal Yadav

April 2015

Thesis Supervisor : Professor Arunava Sen

Thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy

Acknowledgements

First and foremost, I am thankful to my supervisor Professor Arunava Sen. I want to thank him for being the excellent mentor that he is, and his constant support. I feel priviledged to have been his student. He has been extremely patient and also firm; this ensured that I thought about various ideas at length and with focus. He has always been around to offer insight and encouragement. He has emphasized on quality, rigor and depth. His insight has helped me in finding clarity everytime I was lost, his ideas and our many discussions have made this possible. I thank him the most for his enthusiasm and always managing to transfuse it to us. The process of writing this thesis has been exciting and motivating and this has been possible because of Professor Sen. I have learnt enormously from him. I will never be able to thank him enough.

I am very thankful to Professor Debasis Mishra for several very heplful discussions about my research and his comments. He has patiently answered all my questions and has been supportive and cheerful. His jokes have kept us in high spirits.

The theory reading group in ISI organized by Professors Debasis Mishra and Arunava Sen has kept us involved in reading and presenting current theoretical research. Professor Mishra's lectures on mechanism design in the reading group were very detailed and clear. The reading group has been an excellent platform to understand a variety of new ideas. The discussions in the reading group have been extremely helpful in bringing clarity to my understanding.

I am thankful to Professor Bhaskar Dutta for the very clear and interesting introduction to implementation theory in his lectures at the Delhi School of Economics. His lectures were the starting point of my interest in implementation and social choice.

I also want to thank Luis Corchón, Dinko Dimitrov, Francois Maniquet, Shasikanta Nandeibam, Peter Postl, Shigehiro Serizawa and John Weymark for their comments and suggestions. I am thankful to Anirban Bose, Siddharth Chatterjee, Bipul Saurabh, Swaprava Nath and Maneesh Thakur for very helpful discussions.

Mridu Prabal Goswami who is my co-author and friend must be thanked for patiently working with me. I have learnt a lot from him. I also wish to thank Souvik Roy for our many insightful discussions, particularly about Social Choice.

The ISI faculty and admininistration have always offered support, encouragement and help. Professors Satya Das, Bharat Ramaswami, Chetan Ghate, Abhiroop Mukhopadhyay, Tridip Ray, Arup Pal, E Somanathan and others have been wonderful teachers and very encouraging.

ISI has always felt like home and I must thank all the people who made it possible: Tridip *da* for his kindness at all times and being a parent to all of us; Maneesh sir for being my friend; Sharma *ji*, Niranjana ma'am and Subramanium *ji* who made sure our stipends and conference funding always arrived on time; the cooks and all the hostel staff who looked after us without even having to ask and Deepmala for her help with all the paper work. Pawan Gopalakrishnan, Dushyant Kumar, Anuradha Saha, Prachi Singh, Dhritiman Gupta, Parnika

Praleya, Priyanka Kothari, Sanket Patil, Sidharth Chatterjee, Urvashi Jain, Pranabesh have all been an indispensable part of the time that I spent in ISI.

My friends and colleagues in the PhD programme, have been a very important part of the time that I spent in ISI. Anup Pramanik, Abdul Quadir, Mihir Bhattacharya, Madhav Raghavan and Soumendu Sarkar have always been up for all sorts of conversations. We have had numerous discussions about my work and also theirs, which were very helpful and also a source of encouragement in times when progress was slow. We have been partners in sharing our joys and misery, interests and our future plans. This would not have been possible without them.

I am thankful to my family: Papa has shown me how to follow whatever catches my fancy and interest, Ma and my brother Amogh have been unconditionally supportive and their stability has helped me to keep calm. I am very grateful to Nani for her love.

I am especially grateful to Rasika *ji* and Ajay *ji* for their affection and constant support.

I am thankful to Rohit Lamba, Raginee and Soumendu *da* for being my family; Patel for the craziness that we share and our time together. And I want to thank D for simply having met me.

I would like to express my sincere thanks to all those who have helped me both knowingly and unknowingly over the past years.

To Ma

Contents

Chapter 1

Introduction

This thesis comprises of three chapters relating to strategy-proofness and implementation. We provide a brief description of each chapter below.

1.1 A Hurwicz Type Result in a Model with Public Good Production

We consider a two-good model with an arbitrary number of agents. One of the goods is a public good and the otheris a private good. Each agent has an endowment of the private good and the private good can be converted into the public good using a well-behaved production function. A *Social Choice Function (SCF)* associates an allocation with each admissible preference profile. We impose the following requirements on the SCF.

- *Strategy-proofness*: Agent preferences are assumed to be private information and must be elicited. The SCF therefore must be designed to provide agents with dominant-strategy incentives to reveal their private information truthfully.
- *Pareto-efficiency*: The SCF specifies a Pareto-efficient allocation at every preference profile. If this condition is violated, agents will have incentives to re-trade their received allocations ex-post.
- *Individual Rationality*: Agents' must be at least as well-off as they would had they consumed their private good endowment. This is a minimal requirement for agents to participate voluntarily in the mechanism.

We show that these requirements are incompatible with a minimal continuity requirement on the SCF defined over a "small" preference domain.

For our result, we consider a domain $\mathcal{D}$ that consists of all preferences defined by utility functions of the form

$$U(x_i, y; \theta_i) = \theta_i \sqrt{x_i} + y, \ \theta_i > 0.$$

where x_i and y refer to the levels of the private good and the public good respectively.

The domain $\mathcal{D}$ is a restricted domain - it is a single-crossing domain (see Goswami (2013) and Saporiti (2009)). We consider SCF's that satisfy Pareto-efficiency, individual rationality and continuity (defined with respect to the θ_i parameters) over $\mathcal{D}$. However, the SCFs are strategy-proof over a *larger domain*. This domain consists of $\mathcal{D}$ and preferences that are *common concavifications* [1] of those in $\mathcal{D}$ at every consumption bundle. The entire classical domain satisfies this requirement but significantly smaller domains are sufficient. The public good is produced according to a general cost function $c(y)$ that is strictly increasing and weakly convex. According to our result, there does not exist a SCF satisfying strategy-proofness over the extended domain and Pareto-efficiency, individual rationality and continuity over $\mathcal{D}$.

1.2 Selecting Winners with Partially Honest Jurors

We consider the effect of "partially honest" jurors, (along the lines of Dutta and Sen (2012)) in a model of juror decisions developed in Amorós (2010).

We analyze the problem of choosing the w contestants who will win a competition within a group of $n > w$ competitors. All jurors know who the w best contestants are. All of the jurors commonly observe who the w best contestants are, but they may be biased (in favour of or against some contestants). We assume that some of these jurors are partially honest. A partially honest individual has a strict preference for revealing the true state over lying when truth-telling does not lead to a worse outcome (according to preferences in the true state) than that which obtains when lying. The socially optimal rule is to always select the w best contestants, in every possible state of the world. We first look at many person implementation , when the jury consists of at least two partially honest jurors , whose identity is not known to the planner. We find that the socially optimal rule is Nash implementable if for each pair of contestants, there are two jurors who treat the pair in an unbiased manner and one of these jurors is partially honest. However it is not necessary for the planner to know the identity of the jurors who are fair over a given pair. The result shows that the presence of partially honest jurors expands the scope of implementation. We also analyze the problem , when there are only two jurors and consider cases both with and without the assumption of partial honesty.

1.3 Adjacent non-manipulability and Strategy-proofness in Voting Domains

Incentive compatibility is an important question in any model where the agents have private information. Incentive compatibililty guarantees that every agent truthfully reveals his pri-

[1]See Goswami (2013) and Barberà and Jackson (1995).

vate information , irrespective of the announcements made by the other agents. Incentive compatibility assumes that every feasible preference is a candidiate for manipulation. However in many settings, it is plausible and much more convenient for the mechanism designer to consider rules which are immune to candididate manipulations that are "near" or "close" to the true preference of the agent. We are interested in identifying conditions on the domain , which will imply that every rule which is immmune to local manipulation is also incentive compatible (strategy-proof). Sato (2013b) provides a sufficient condition for the equivalence and a weaker necessary condition. Our main result identifies a weaker sufficient condition for equivalence (than that of Sato (2013b)).

Chapter 2

A Hurwicz Type Result in a Model with Public Good Production

2.1 Introduction

A classic result in the theory of incentive compatibility is Hurwicz (1972). The paper considered a two-good, two-agent exchange economy and showed the non-existence of strategy-proof, Pareto-efficient and individually rational social choice functions in this environment. There is a large literature extending and refining this result for arbitrary exchange economies.[1] In this paper, we consider the same issue in the context of production economies.

We consider a two-good model with an arbitrary number of agents. One of the goods is a public good and the other, a private good. Each agent has an endowment of the private good and the private good can be converted into the public good using a well-behaved production function.

A *Social Choice Function (SCF)* associates an allocation with each admissible preference profile. Some standard requirements on SCF's are imposed.

- *Strategy-proofness*: Agent preferences are assumed to be private information and must be elicited. The SCF therefore must be designed to provide agents with dominant-strategy incentives to reveal their private information truthfully.

- *Pareto-efficiency*: The SCF specifies a Pareto-efficient allocation at every preference profile. If this condition is violated, agents will have incentives to re-trade their received allocations ex-post.

- *Individual Rationality*: Agents' must be at least as well-off as they would had they consumed their private good endowment. This is a minimal requirement for agents to participate voluntarily in the mechanism.

[1] See Serizawa (2002), Serizawa and Weymark (2003), Nicoló (2004), Serizawa (2006), Zhou (1991), Goswami et al. (2014), Barberà and Jackson (1995) and Hashimoto (2008).

We show that these requirements are incompatible with a minimal continuity requirement on the SCF defined over a "small" preference domain.

For our result, we consider a domain $\mathcal{D}$ that consists of all preferences defined by utility functions of the form

$$U(x_i, y; \theta_i) = \theta_i \sqrt{x_i} + y,\ \theta_i > 0.$$

where x_i and y refer to the levels of the private good and the public good respectively.

The domain $\mathcal{D}$ is a restricted domain - it is a single-crossing domain (see Goswami (2013) and Saporiti (2009)). We consider SCF's that satisfy Pareto-efficiency, individual rationality and continuity (defined with respect to the θ_i parameters) over $\mathcal{D}$. However, the SCFs are strategy-proof over a *larger domain*. This domain consists of $\mathcal{D}$ and preferences that are *common concavifications* [2] of those in $\mathcal{D}$ at every consumption bundle. The entire classical domain satisfies this requirement but significantly smaller domains are sufficient (details can be found in Section 2.1). The public good is produced according to a general cost function $c(y)$ that is strictly increasing and weakly convex. According to our result, there does not exist a SCF satisfying strategy-proofness over the extended domain and Pareto-efficiency, individual rationality and continuity over $\mathcal{D}$.

Several papers have examined the relationship between our axioms in models with public good production. These papers differ from ours in their choice of axioms, the domain and the nature of the cost function. The main contribution of our paper is that we establish our impossibility result over a specific "narrow" domain with a general cost function using standard axioms. We briefly outline the relationship of our result with those that already exist.

Hurwicz and Walker (1990) demonstrate the incompatibility of strategy-proofness and Pareto-efficiency for interior allocations in the standard quasi-linear domain i.e. utility functions of the form $v_i(y, \theta_i) + x_i$.[3] Beviá and Corchón (1995) prove an impossibility result in the same model with a linear production technology. They require the individual rationality axiom, but show manipulability at all profiles. The domain must be chosen carefully because strategy-proofness and Pareto-efficiency are compatible in specific domains - for instance in the case where the function $v_i(y, \theta_i)$ has a quadratic form (see Tian (1996) for a generalization). Showing that x_i allocations in the quasi-linear domain where Pareto-efficiency and strategy-proofness are compatible (such as the quadratic case) cannot be designed to satisfy to individual rationality, is not straight forward. This is because these allocations are not unique as pointed out in Tian (1996).

Our model has very different structural features from the quasi-linear model discussed above. Unlike the quasi-linear case, Pareto-efficiency no longer identifies a unique level of public good. Our result also extends to the case where the production technology is not

[2]See Goswami (2013) and Barberà and Jackson (1995).

[3]In their model, although production of the public good does not appear explicitly, it can be embedded without loss of generality in a model with endowments and a linear production technology.

linear. Overall our methods are very different from those usually employed in the quasi-linear case.

Corchón and Rueda-Llano (2008) use a domain similar to our basic domain. They work on a domain consisting of preferences representable by a additively seperable utility function satisfying a single-crossing property. It restricts attention to *twice-continuously differentiable* SCF's and demonstrates non-existence of a two-person SCF satisfying Pareto-efficiency, strategy-proofness and non-dictatorship. The generalization to an arbitrary number of agents requires individual rationality and an additional axiom called *weak regularity*. Weak regularity is a technical property that is satisfied if there is at least one agent in the economy for whom telling the truth is the unique best dominant strategy.

Serizawa (1996) considers the domain of all continuous, strictly monotone and strictly convex preferences in a single public good-single private good model. It characterizes a particular public good provision/cost sharing rule as the unique rule which is strategy-proof, *non-exploitative*, individually rational and non-bossy. A SCF is non-exploitative if no agent receives a consumption of the private good larger than her endowment. A SCF is non-bossy if no agent can affect the bundle consumed by any other agent without affecting the bundle consumed by him. This rule violates Pareto-efficiency. An immediate consequence is that there does not exist a SCF satisfying strategy-proofness, non-exploitation, individual rationality, non-bossiness and Pareto-efficiency. Deb and Ohseto (1999) show that a strategy-proof, individually rational SCF satisfying no-exploitation on the full class of preferences also satisfies non-bossiness. Therefore the non-bossiness requirement in the impossibility result can be dropped, i.e. there does not exist a SCF satisfying strategy-proofness, non-exploitation, individual rationality and Pareto-efficiency.[4]

Our result differs from both Corchón and Rueda-Llano (2008) and Serizawa (1996) in important ways. Our continuity assumption is clearly weaker than the differentiability requirement in the former and we do not make the weak regularity assumption. On the other hand, our domain includes non-single crossing preferences. Our result is also not comparable to the impossibility result in Serizawa (1996) since it uses an additional continuity assumption while not requiring non- exploitativeness nor non-bossiness. We note that our result does not make a non-bossiness assumption, an assumption that is pervasive in the literature on allocation models with at least three agents.[5]

An alternative approach to the problem can be formulated using ideas in Jackson (2003). Assume that the cost of producing the public good is shared equally amongst the agents. The utility function $U(x_i, y; \theta_i)$ can then be written as a function of y alone, for every θ_i i.e. as $v_i(y; \theta_i)$. Standard assumptions on $U_i(x_i, y, \theta_i)$ ensure that $v_i(y; \theta_i)$ is *single-peaked*.

[4]The no-exploitation condition can be regarded as weak. However our model and assumptions allow for the possibility of its violation. On the other hand, we impose a continuity assumption albeit on a specific sub-domain. Our results are therefore independent of Serizawa (1996) and Deb and Ohseto (1999).

[5]See Satterthwaite and Sonnenschein (1981), Serizawa (1996), Barberà and Jackson (1995) and Goswami et al. (2014).

Denote the peak of $v_i(y;\theta_i)$ by $\hat{y}(\theta_i)$. A sufficiently rich class of $U_i(x_i, y; \theta_i)$ functions will generate all single-peaked preferences on y. Well known results on strategy-proof rules on single-peaked domains (Moulin (1980), Weymark (2011)) can be applied to show that that the public good production rule will be a median rule with phantom voters - a precise result is stated as Theorem 3 in Jackson (2003). The only rule satisfying individual rationality in addition to strategy-proofness, is the *minimum demand rule*. This rule picks $\min_i \hat{y}(\theta_i)$ amongst the agents. However this rule is not Pareto efficient - for example, a change in the function $v_i(y;\theta_i)$ for all i that leaves $\hat{y}(\theta_i)$ unchanged will keep the chosen level of public good unchanged but will typically affect its Pareto efficient level. In the paper, we consider a more general model where the cost shares are not fixed and can depend on the reported vector of types. This also implies that arguments using single-peakedness cannot be extended easily. As the cost sharing parameter changes, the $v_i(.)$ function itself changes which will typically lead to complications. However the idea of reducing the problem to one of choosing the level of the public good in a single-peaked domain (assuming a fixed cost share) serves as a useful background to our result.

2.2 The Model

We consider an economy with a set of agents $I = \{1, 2, .., n\}$. There is a single public good that can be produced from a private good according to a convex cost function. We let $y \in \Re_+$ denote the amount of public good and $x_i \in \Re_+$, the consumption of private good by agent i. Agent i has an initial endowment of the private good $\omega_i > 0$. The aggregate endowment is $\omega = \sum_{i \in I} \omega_i$.

The cost of producing the public good is given by the cost function $c : \Re_+ \to \Re_+$. We assume that this function is twice-continuously differentiable satisfying the following conditions (i) $\infty > c'(y) \geq 0$ for $y \geq 0$ with $c'(y) > 0$ whenever $y > 0$ and (ii) $c''(y) \geq 0$ for all $y \geq 0$.

The set of feasible allocations is

$$A \equiv \left\{ (y, x_1, x_2, .., x_n) : y \in \Re_+, \; x_i \in \Re_+ \; \forall i, \; \sum_{i \in I} x_i + c(y) \leq \omega \right\}.$$

Let y^* be the maximum amount of public good that can be produced given the aggregate endowment of the private good is ω, i.e. $y^* = c^{-1}(\omega)$.

2.2.1 Preferences

We assume that each agent i has a preference ordering over private-good, public-good bundles. Such a preference will be denoted by R_i. For any pair of private-good, public-good bundles (x_i, y) and $(x_i^{'}, y^{'})$, if $(x_i, y) \; R_i \; (x_i^{'}, y^{'})$ then (x_i, y) is weakly preferred to $(x_i^{'}, y^{'})$. The asymmetric and symmetric components will be denoted by P_i and I_i respectively. We let

$\mathcal{R}$ denote the set of *all* preference orderings that are (a) continuous, (b) strictly monotone in both public and private good and (c) strictly convex. We refer to such preferences as *classical* preferences. A profile $R = (R_1, \ldots, R_n)$ is an n-tuple of preference orderings, one for each agent. For any $S \subseteq N$ and profile R, $R_{N \setminus S}$ is the profile of $N - S$ agents where the preference orderings of agents in S are deleted. In the special case where $S = \{i\}$, we write R_{-i} for $R_{N \setminus S}$. The set of all profiles is $\mathcal{R}^n$.

Our starting point is the domain $\mathcal{D}$ of preferences that can be represented by utility functions of the form $u_i(x_i, y; \theta_i) = \theta_i \sqrt{x_i} + y$ where $\theta_i > 0$. These preferences are parametrized by the positive real number θ_i.[6] We shall denote the indifference curve of preference θ_i through point allocation (x_i, y) by $IC(\theta_i, (x_i, y))$. A preference profile in $\mathcal{D}$ can be represented by an n-tuple $\theta = (\theta_1, \ldots, \theta_n)$. Let $S \subseteq N$, $\theta_S = (\theta_i)_{i \in S}$ and $R \in \mathcal{R}^n$, $(\theta_S, R_{N \setminus S})$ denotes the profile where agents in S and $N \setminus S$ have preference orderings in $\mathcal{D}$ and $\mathcal{R}$.

We will operate on a strictly larger domain of $\mathcal{D}$ which we call $\mathcal{D}^{cc}$. The larger domain consists of preferences in $\mathcal{D}$ as well as *common concavifications* of preferences in $\mathcal{D}$. We first define concavification.

Let (x_i, y) be a private-good,public-good pair and R_i be a preference of agent i. Then $UC(R_i, (x_i, y))$ is the set of commodity bundles that are at least as good as (x_i, y) according to R_i.

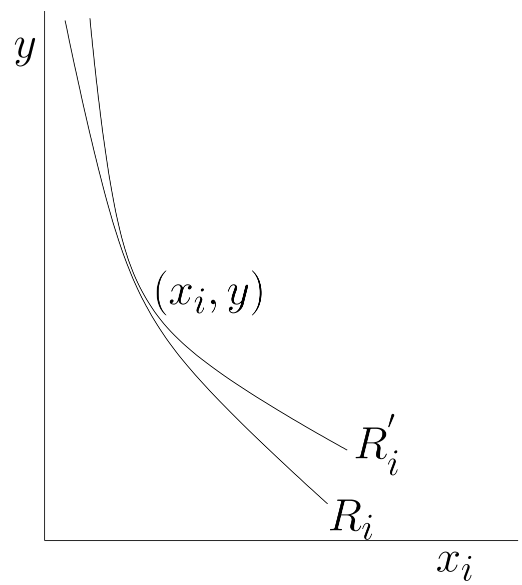

Figure 2.1: Concavification

Definition 1 *Let R_i be a preference ordering and let $(x_i, y) \in \Re^2_+$ be a private-good,public-good pair. The preference ordering R_i' is a concavification of R_i at (x_i, y) if*

[6] These preferences have the *single-crossing* property; see Goswami et al. (2014) and Goswami (2013). They clearly satisfy properties (a), (b) and (c) above and therefore belong to $\mathcal{R}$.

(i) $UC(R_i', (x_i, y)) \subseteq UC(R_i, (x_i, y))$ *and*

(ii) $(x_i', y') \in UC(R_i', (x_i, y))$ *and* $(x_i', y') \neq (x_i, y) \implies (x_i', y')\ P_i\ (x_i, y)$.

This is illustrated in Figure 2.1. Common concavification requires the existence of a preference that simulataneously concavifies two specifically chosen preferences in $\mathcal{D}$ for two (x_i, y) bundles chosen in a particular way. Formally,

Definition 2 *Let* $(x_i, y), (x_i', y') \in \Re_+^2$. *The bundles* (x_i, y) *and* (x_i', y') *are denoted by* a *and* b *respectively and* $L(a, b)$ *denotes the line joining* a *and* b. *Let* θ_i, θ_i' *be such that*

(i) $\theta_i\sqrt{x_i'} + y' > \theta_i\sqrt{x_i} + y$

(ii) $\theta_i'\sqrt{x_i'} + y' > \theta_i'\sqrt{x_i} + y$.

(iii) $\theta_i' > \theta_i$.

(iv) If $L(a, b)$ *is downward sloping, then we require that* $IC(\theta_i, b)$ *cuts* $L(a, b)$ *from above at point* b. [7]

Then $R_i \in \mathcal{R}$ *is a common concavification if* R_i *is a concavification of* θ_i *at* (x_i, y) *and a concavification of* θ_i' *at* (x_i', y').

This is illustrated in Figure 2.2.

The Appendix provides the details for such a construction. The additional preferences are classical. They remain "narrow" in the sense that they are strict sub-class of the set of preferences that are seperable in the public and private good.

2.2.2 Social Choice Functions

In this subsection, we recall some basic definitions.

A *Social Choice Function* or SCF is a map $F : \mathcal{R}^n \to A$. We let $F^* : \mathcal{D}^n \to A$ denote the restriction of F to the domain $\mathcal{D}^n$. For any profile R, $F_i(R) = (x_i(R), y(R))$ where $i \in I$. We will write $x_i(R)$ and $y(R)$ respectively as the private good allocated to agent i and the public good produced at profile R by F. Thus, F comprises $(n+1)$ functions, $x_i : \mathcal{R}^n \to [0, \omega]$, $i = 1, \ldots, n$ and $y : \mathcal{R}^n \to [0, y^*]$. For a general profile $(\theta_S, R_{N\setminus S})$ where $S \subseteq N$, $x_i(\theta_S, R_{N\setminus S})$ and $y(\theta_S, R_{N\setminus S})$ will denote the private good allocated to agent i and the public good produced respectively. Note that in view of our earlier remarks, $F^* : \Re_{++}^n \to A$.

We now describe some properties of SCF's.

[7] The absolute value of the slope of $IC(\theta_i, b)$ at point b is strictly greater than the absolute value of the slope of $L(a, b)$.

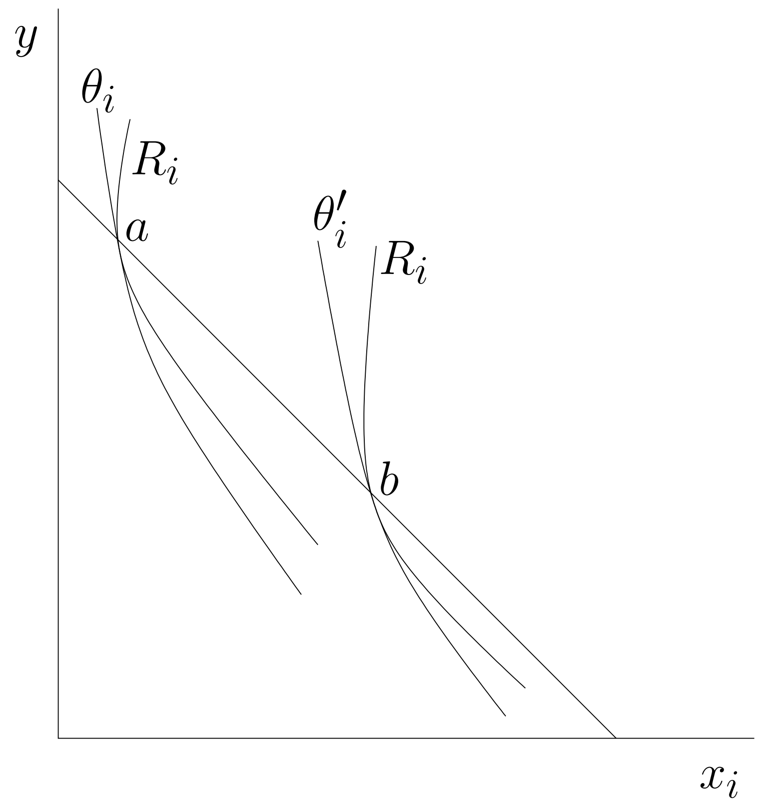

Figure 2.2: Common Concavification

DEFINITION **3** *An SCF F is manipulable on $[\mathcal{D}^{cc}]^n$ if there exists agent i and a profile R such that*

$$F(R_i', R_{-i})\ P_i\ F(R_i, R_{-i}).$$

It is strategy-proof (SP) if it is not manipulable by any agent at any profile.

Strategy-Proofness is the standard notion of Dominant Strategy Incentive Compatibility. If an SCF is Strategy-Proof, no agent can be strictly better off by misrepresenting his preferences irrespective of the announcements of the other agents. Note that strategy-proofness is defined on the sub-domain $[\mathcal{D}^{cc}]^n$.

DEFINITION **4** *An SCF F is Pareto-efficient on the restricted domain $\mathcal{D}^n$ (PERD) if F^* is Pareto-efficient i.e. there does not exist $\theta \in \Re^n_{++}$ and $(y, x_1, ..., x_n) \in A$ such that*

$$\theta_i\sqrt{x_i} + y \geq \theta_i\sqrt{x_i(\theta)} + y(\theta)$$

for all i with at least one strict inequality.

We require F to be Pareto-efficient *only* on the sub-domain $\mathcal{D}^n$. Therefore PERD is a weaker requirement than standard Pareto-Efficiency which would apply to the whole domain. A similar remark holds for our notion of Individual Rationality which we define below.

Definition 5 *An SCF F is Individually Rational on the restricted domain $\mathcal{D}^n$ (IRED) if F^* is Individually Rational i.e.*

$$\theta_i\sqrt{x_i(\theta)} + y(\theta) \geq \theta_i\sqrt{\omega_i}.$$

for all i and for all θ.

The Individual Rationality axiom ensures that no agent is strictly better-off by rejecting the allocation prescribed by the mechanism and consuming only her endowment (of private good). The implicit assumption is that if an agent chooses not to participate in the mechanism, she will be excluded from public good consumption. This is a weaker requirement than the one that would hold if the agent was allowed to free-ride on the public good supplied by the contribution of other agents. This is also the approach of Corchón and Rueda-Llano (2008). Serizawa (1996) follows an alternative and more general approach. In his model, Serizawa assumes that all agents have access to a personalized public good producing technology which can be operated by a contribution from the agent's own private good endowment. In the special case where this private technology is the "null technology", Serizawa's individual rationality condition reduces to ours.

Definition 6 *A SCF F is restricted domain continuous (RDC) if F^* is continuous.*

Since $F^* : \Re^n_{++} \to A$, there are no technical issues in defining continuity.

2.3 The Result

Our result is the following.

Theorem 1 *There does not exist a SCF satisfying Strategy-Proofness, PERD, IRED and RDC.*

We provide a brief description of the proof strategy. We prove the result by way of contradiction, ie. we assume that there exists a SCF which satisfies all the axioms. The first step is to show that if the SCF satisfies IRED, then for any arbitrary profile, there will exist an agent, say i whose private good allocation is strictly positive. Strategy-proofness implies that both $x_i(\cdot)$ and $y(\cdot)$ are monotonic in θ_i. A key step in the argument is to show that $x_j(\cdot)$ is also monotonic (non decreasing) in θ_i. This follows from Strategy proofness, RDC and the fact that the public good allocation affects the utility of all agents. The next step is to establish that as the type of agent i becomes arbitrarily large, the public good allocation tends to zero. This fact implies that the SCF cannot satisfy PERD, leading to a contradiction.

2.4 Proof

The proof uses the following key lemmas.

Lemma 1 (PERD Characterization) *If F satisfies PERD, there exists $\alpha(\theta) \in \mathbb{R}^n_+$ with $(0,\dots,0) \neq \alpha(\theta) \leq (1,\dots,1)$ satisfying*

$$\sum_{i\in N} \alpha_i(\theta) \frac{2\sqrt{x_i(\theta)}}{\theta_i} \leq c'(y(\theta)) \tag{2.1}$$

$$\left[\sum_{i\in N} \alpha_i(\theta) \frac{2\sqrt{x_i(\theta)}}{\theta_i} - c'(y(\theta))\right] y(\theta) = 0. \tag{2.2}$$

$$(1-\alpha_i(\theta))x_i(\theta) = 0 \textit{ for all } i \in N. \tag{2.3}$$

These conditions are a special case of the conditions derived for a more general model in Campbell and Truchon (1988).

If a Pareto-efficient allocation satisfies $x_i(\theta) > 0$ for all i, then $\alpha_i(\theta) = 1$ for all i. If a Pareto efficient allocation satisfies $y(\theta) > 0$, then Equation 2.1 holds with equality.

Lemma 2 (Monotonicity) *Let F be SP. Pick $i \in I$ and $R_{-i} \in [\mathcal{D}^{cc}]^{n-1}$. Then*

(i) $x_i(\theta_i, R_{-i})$ is non-decreasing in θ_i.

(ii) $y(\theta_i, \theta_{-i})$ is non-increasing in θ_i.

Proof: Consider an agent i with preference θ_i and θ_i' with $\theta_i' > \theta_i$.

From strategy-proofness

$$\theta_i\sqrt{x_i(\theta_i, R_{-i})} + y(\theta_i, R_{-i}) \geq \theta_i\sqrt{x_i(\theta_i', R_{-i})} + y(\theta_i', R_{-i}) \tag{2.4}$$

and

$$\theta_i'\sqrt{x_i(\theta_i', R_{-i})} + y(\theta_i', R_{-i}) \geq \theta_i'\sqrt{x_i(\theta_i, R_{-i})} + y(\theta_i, R_{-i}). \tag{2.5}$$

Adding Equations 2.4 and 2.5 and rearranging, we have

$$(\theta_i' - \theta_i)\left(\sqrt{x_i(\theta_i', R_{-i})} - \sqrt{x_i(\theta_i, R_{-i})}\right) \geq 0. \tag{2.6}$$

Equation 2.6 establishes (i).

Rearranging Equation 2.4 also gives

$$y(\theta_i, R_{-i}) - y(\theta_i', R_{-i}) \geq \theta_i\left(\sqrt{x_i(\theta_i', R_{-i})} - \sqrt{x_i(\theta_i, R_{-i})}\right). \tag{2.7}$$

Equation 2.7 along with (i) implies (ii). ■

Lemma **3 (Common Concavification Lemma)** *Let* $(x_i, y), (x_i^{'}, y^{'}) \in \Re^2_+$. *The bundles* (x_i, y) *and* $(x_i^{'}, y^{'})$ *are denoted by* a *and* b *respectively and* $L(a, b)$ *denotes the line joining* a *and* b. *Let* $\theta_i, \theta_i^{'}$ *be such that*

(i) $\theta_i\sqrt{x_i^{'}} + y^{'} > \theta_i\sqrt{x_i} + y$

(ii) $\theta_i^{'}\sqrt{x_i^{'}} + y^{'} > \theta_i^{'}\sqrt{x_i} + y$.

(iii) $\theta_i^{'} > \theta_i$.

(iv) If $L(a, b)$ *is downward sloping, then we require that* $IC(\theta_i, b)$ *cuts* $L(a, b)$ *from above at point* b. [8]

Then there exists $R_i \in \mathcal{R}$ *such that* R_i *is a concavification of* θ_i *at* (x_i, y) *and a concavification of* $\theta_i^{'}$ *at* $(x_i^{'}, y^{'})$.

A formal proof of Lemma 3 is contained in the Appendix. We note that (iv) in the choice of a and b is a technical requirement that considerably simplifies the construction of the common concavification. The Lemma is true without the requirement. However the circumstances where the concavification lemma is used in our proof satisfy (iv). So the simplified construction suffices for our purpose. Requirements (i) and (ii) are necessary for common concavification.

We now begin the proof of the theorem.

Proof: Suppose the theorem is false i.e. F is a SCF satisfying SP, PERD, IERD and RDC. Recall that F^* is the restriction of F to the domain $\mathcal{D}^n$.

For the rest of the proof till the end of Claim 3, we fix an arbitrary agent i and $\hat{\theta}_{-i}$. We claim that there exists $\hat{\theta}_i$ such that $x_i(\hat{\theta}_i, \hat{\theta}_{-i}) > 0$.

If the claim is false, then IRED for agent i implies

$$y(\theta_i, \hat{\theta}_{-i}) \geq \theta_i\sqrt{\omega_i} \qquad \text{for all } \theta_i. \tag{2.8}$$

The LHS of (2.8) is bounded above by y^*. On the other hand, the RHS can be made arbitrarily large by choosing θ_i large enough, leading to a contradiction.

In view of Lemma 2, $x_i(\theta_i, \hat{\theta}_{-i})$ and $y(\theta_i, \hat{\theta}_{-i})$, are non-decreasing and non-increasing in θ_i respectively. We will show the following:

CLAIM 1: For all $j \neq i$, $x_j(\theta_i, \hat{\theta}_{-i})$ is non-decreasing in θ_i.

Suppose Claim 1 is false i.e. there exists $\theta_i^{\prime}$ and $\theta_i^{\prime\prime}$ with $\theta_i^{\prime} > \theta_i^{\prime\prime}$ and $j \neq i$ such that and $x_j(\theta_i^{\prime}, \hat{\theta}_{-i}) < x_j(\theta_i^{\prime\prime}, \hat{\theta}_{-i})$. Note that Lemma 2 implies $y(\theta_i^{\prime}, \hat{\theta}_{-i}) \leq y(\theta_i^{\prime\prime}, \hat{\theta}_{-i})$.

The functions $x_j = x_j(\theta_i, \hat{\theta}_{-i})$ and $y = y(\theta_i, \hat{\theta}_{-i})$ are continuous in θ_i since F satisfies RDC. Therefore, θ_i parametrizes a curve in $\Re^2_+$ according to $\theta_i \mapsto ((x_j(\theta_i, \hat{\theta}_{-i}), y(\theta_i, \hat{\theta}_{-i}))$.

[8] The absolute value of the slope of $IC(\theta_i, b)$ at point b is strictly greater than the absolute value of the slope of $L(a, b)$.

(See Rudin (1976), page 131). Consider the section of the curve as θ_i varies from θ_i'' to θ_i', labelled A in Figure 2.3. The curve must pass through the points a and b. By Lemma 2 the curve must lie below the horizontal line $y = y(\theta_i'', \hat{\theta}_{-i})$ and above $y = y(\theta_i', \hat{\theta}_{-i})$. These facts imply the existence of $\theta_i^* \in (\theta_i'', \theta_i']$ such that $x_j(\theta_i, \hat{\theta}_{-i}) > 0$ and $x_j(\theta_i, \hat{\theta}_{-i}) \geq x_j(\theta_i', \hat{\theta}_{-i})$ for all $\theta_i \in (\theta_i'', \theta_i^*) \subseteq (\theta_i'', \theta_i')$. Figure 2.3 shows θ_i^* for alternative paths A.

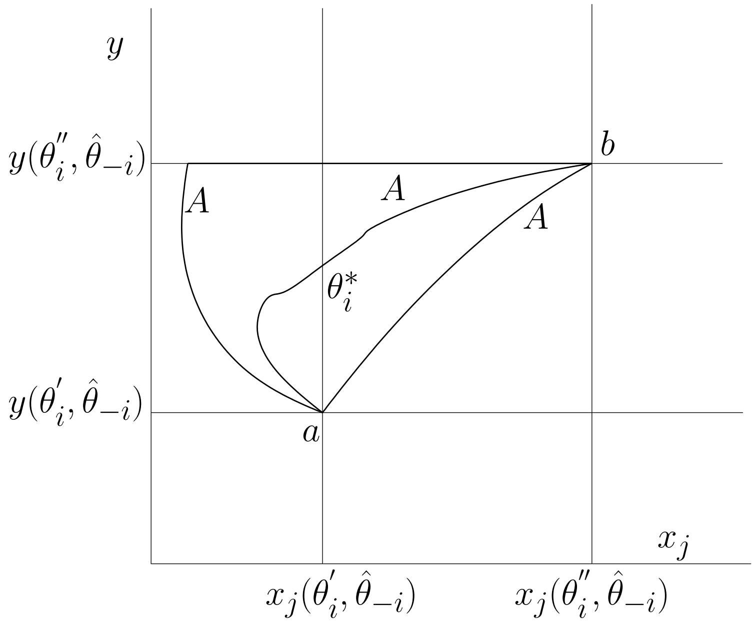

Figure 2.3: Alternative path A's

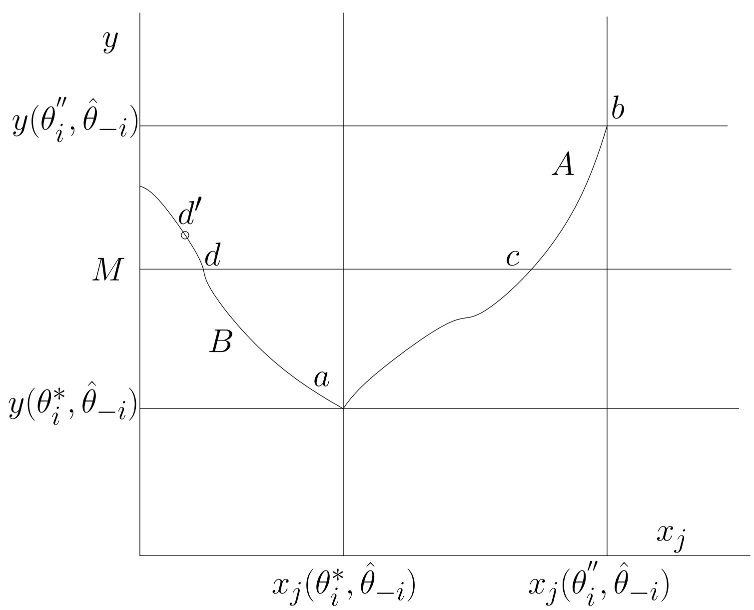

Figure 2.4: Monotonicity of x_j in θ_i

Let B be the path parametrized by θ_j according to $\theta_j \to ((x_j(\theta_i^*, \theta_j, \hat{\theta}_{-ij}), y(\theta_i^*, \theta_j, \hat{\theta}_{-i,j}))$ for $\theta_j \leq \hat{\theta}_j$. Let a denote the point $(x_j(\theta_i^*, \hat{\theta}_j, \hat{\theta}_{-i,j}), y(\theta_i^*, \hat{\theta}_j, \hat{\theta}_{-i,j}))$ (Figure 2.4). Suppose B is degenerate at a i.e. the path B consists of only point a. This would imply $y(\theta_i^*, \theta_j, \hat{\theta}_{-i,j}) = y(\theta_i^*, \hat{\theta}_j, \hat{\theta}_{-i,j}$ for all $\theta_j < \hat{\theta}_j$. By SP for agent j, $x_j(\theta_i^*, \theta_j, \hat{\theta}_{-i,j}) = x_j(\theta_i^*, \hat{\theta}_j, \hat{\theta}_{-i,j})$ for all $\theta_j < \hat{\theta}_j$.

By PERD,

$$\frac{2\sqrt{x_j(\theta_i^*, \theta_j, \hat{\theta}_{-i,j})}}{\theta_j} + \sum_{k \neq j} \alpha_k(\theta_i^*, \theta_j, \hat{\theta}_{-i,j}) \frac{2\sqrt{x_j(\theta_i^*, \theta_j, \hat{\theta}_{-i,j})}}{\theta_k} = c'(y(\theta_i^*, \theta_j, \hat{\theta}_{-i,j})). \tag{2.9}$$

The LHS of Equation 2.9 can be made arbitratrily large by choosing θ_j sufficiently small. On the other hand, the RHS is a finite constant. Therefore B cannot be degenerate. Let $\lim_{\theta_j \to 0} x_j(\theta_i^*, \theta_j, \hat{\theta}_{-i,j}) = \bar{x}_j$ and $\lim_{\theta_j \to 0} y(\theta_i^*, \theta_j, \hat{\theta}_{-i,j}) = \bar{y}$. We can therefore infer that $\bar{x} < x_j(\theta_i^*, \hat{\theta}_{-i})$ and $\bar{y} > y(\theta_i^*, \hat{\theta}_{-i})$.

Pick M such that $y(\theta_i^*, \hat{\theta}_{-i}) < M < y(\theta_i^{''}, \hat{\theta}_{-i})$. Let $\theta_j^{'}$ and $\tilde{\theta}_i$ be such that $y(\theta_i^*, \theta_j^{'}, \hat{\theta}_{-i,j}) = y(\tilde{\theta}_i, \hat{\theta}_j, \hat{\theta}_{-i,j}) = M$. The existence of $\theta_j^{'}$ and $\tilde{\theta}_i$ is guaranteed by our earlier arguments (See Figure 2.4 and points d and c).

We will find a common concavification for agent j at point c and a point d' on B, arbitratrily close to d such that the level of public good at d' is strictly greater than M. In order to apply Lemma 3, we need to ensure that c is strictly preferred to d' at both $\hat{\theta}_j$ and $\theta_j^{''}$ which corresponds to the preference for which the allocation for agent j at $(\theta_i^*, \theta_j^{''}, \hat{\theta}_{-i,j})$ is d'. Note that condition (iii) in Lemma 3 is satisfied for any point d' such that d' is distinct from a. We also need to satisfy condition (iv) in Lemma 3. We claim that all these properties can be satisfied by a suitable choice of d'.

Due to the monotonicity of preferences, c is strictly preferred to d at $\hat{\theta}_j$. Since preferences are also continuous, there is a neighbourhood of d where all allocations will be strictly worse than c under $\hat{\theta}_j$. Formally, let $LC(\hat{\theta}_j, c)$ denote the lower contour set of allocation c under preference $\hat{\theta}_j$. Then $d \in \text{Int } LC(\hat{\theta}_j, c)$.[9] Thus there exists a neighbourhood of d, $B_\delta(d)$ such that $B_\delta(d) \subsetneq \text{Int } LC(\hat{\theta}_j, c)$. The point d' can be chosen from this neighbourhood to satisfy condition (ii) of Lemma 3.

As discussed above, we need to ensure that c is strictly preferred to d' at $\theta_j^{''}$ where d' corresponds to the preference profile $(\theta_i^*, \theta_j^{''}, \hat{\theta}_{-i,j})$. In order to find such a d', we prove the following claim.

Let $(x_j^{'}, y^{'}) \in \text{Int } LC(\theta_j^*, (x_j^*, y^*))$. Then there exists a neighbourhood of θ_j^*, $(\theta_j^* - \epsilon, \theta_j^* + \epsilon)$ and a neighbourhood of $(x_j^{'}, y^{'})$, $B_\delta((x_j^{'}, y^{'}))$ such that $B_\delta((x_j^{'}, y^{'})) \subsetneq \text{Int } LC(\theta_j, (x_j^*, y^*)$ for all $\theta_j \in (\theta_j^* - \epsilon, \theta_j^* + \epsilon)$.

We have $\theta_j^* \sqrt{x_j^{'}} + y^{'} < \theta_j^* \sqrt{x_j^*} + y^*$. This means $\theta_j^* \big[\sqrt{x_j^{'}} - \sqrt{x_j^*}\big] + y^{'} - y^* < 0$. We define $g(\theta_j, x_j, y) = \theta_j \big[\sqrt{x_j} - \sqrt{x_j^*}\big] + y - y^*$. Hence, we have $g(\theta_j^*, x_j^{'}, y^{'}) < 0$. Note that

[9] Int S denotes the interior of set S.

$g(.)$ is a continuous function on $\Re^3_{++}$. Thus by continuity of $g(.)$, we have a neighbourhood of $(\theta_j^*, x_j^{'}, y^{'})$, $B_\delta(\theta_j^*, x_j^{'}, y^{'})$ such that for all $(\theta_j, x_j, y) \in B_\delta(\theta_j^*, x_j^{'}, y^{'})$, $g(\theta_j, x_j, y) < 0$. Hence the above claim follows.

Due to monotonicity of preferences c is strictly preferred to d at $\theta_j^{'}$. Thus by the above claim, there exists a neigbourhood of d such that all points in this neighbourhood are strictly worse than c at any θ_j which belongs to the interval $(\theta_j^{'} - \epsilon, \theta_j^{'} + \epsilon)$. Thus any point d' in this neighbourhood will satisfy condition (i) of Lemma 3. Since $x_j(\theta_j, \theta_{-j})$ and $y(\theta_j, \theta_{-j})$ are both monotone in θ_j, the curve B is differentiable almost everywhere. Strategy proofness of agent j implies that at any differentiable point e on Path B, the slope of $IC(\theta_j^*, e)$ (where e corresponds to the profile $(\theta_i^*, \theta_j^*, \hat{\theta}_{-i,j})$) is equal to the derivative of Path B at point e. By the single crossing property, θ_j^* is unique. Thus now we can choose d' (which uniquely corresponds to $\theta_j^{''}$) on path B in the appropriate neighbourhood such that conditions (i) and (ii) of Lemma 3 are satisfied. We can therefore find a $\theta_j^{''}$ close to $\theta_j^{'}$ such that d' is on path B and satisfies conditions (i) and (ii) of Lemma 3.

Condition (iv) in Lemma 3 requires $IC(\theta_j^{''}, c)$ cuts $L(d', c)$ from above at c. [10] However the existence of d' and $\theta_j^{''}$ as described above does not guarantee that $IC(\theta_j^{''}, c)$ will cut $L(d', c)$ from above.

We will now argue that our choice of d' satisfies condition (iv) as well. Since the SCF is continuous, from our arguments above we can infer that for every $\epsilon > 0$, there exists $\theta_j^\epsilon \in (\theta_j^{'} - \epsilon, \theta_j^{'})$ and point d^ϵ on Path B close to d such that conditions (i) and (ii) of Lemma 3 are satisfied.

In the neighbourhood that we defined above to satisfy requirements (i) and (ii) of Lemma 3, we can construct a strictly decreasing sequence $\{\epsilon^k\}$ converging to 0. Note that every ϵ^k corresponds to a point d^k on Path B and θ_j^k. Also by construction, $\theta_j^k < \theta_j^{k+1} < \theta_j^{'}$. The sequence $\{\theta_j^k\}$ converges to $\theta_j^{'}$.

We define the function $h(\theta_j^k)$ as follows:

$$h(\theta_j^k) = -\frac{y(\theta_i^*, \theta_j^k, \hat{\theta}_{-i,j}) - y(\tilde{\theta}_i, \hat{\theta}_j, \hat{\theta}_{-i,j})}{x_j(\theta_i^*, \theta_j^k, \hat{\theta}_{-i,j}) - x_j(\tilde{\theta}_i, \hat{\theta}_j, \hat{\theta}_{-i,j})}.$$

The function $h(.)$ is the slope (absolute value) of the line $L(d^k, c)$. Note that $h(.)$ is continuous in θ_j due to the continuity of the SCF. Thus $h(.)$ tends to 0 as θ_j^k converges to $\theta_j^{'}$.

The slope (absolute value) of $IC(\theta_j^k, c)$ at point c is defined by the function $f(\theta_j^k)$,

$$f(\theta_j^k) = \frac{\theta_j^k}{2\sqrt{x_j(\tilde{\theta}_i, \hat{\theta}_j, \hat{\theta}_{-i,j})}}.$$

[10] $L(d', c)$ is downward sloping since d' is chosen such that the level of public good at d' is strictly greater than M which is the level of public good at c. Also the level of private good at d' is strictly less than the level of private good at c.

The limit of the function $f(\theta_j^k)$ as θ_j^k tends to θ_j' is $\frac{\theta_j'}{2\sqrt{x_j(\tilde{\theta}_i,\hat{\theta}_j,\hat{\theta}_{-i,j})}}$ which is strictly positive. Also, the function $f(\theta_j^k)$ is strictly increasing.

Thus there exists θ_j^K such that

$$f(\theta_j^K) > h(\theta_j^K).$$

Hence we obtain $\theta_j'' = \theta_j^K$ and $d' = d^K$ which satisfies all the conditions of Lemma 3. Let $(x_j(c), y(c))$ and $(x_j(d'), y(d'))$ denote the private-good to j, public good pairs corresponding to c and d' respectively.

Applying Lemma 3, there exists an R_j, which concavifies θ_j'' at d' and $\hat{\theta}_j$ at c (see Figure 2.4). By SP, $(x_j(\theta_i^*, R_j, \hat{\theta}_{-i,j}), y(\theta_i^*, R_j, \hat{\theta}_{-i,j})) = (x_j(d'), y(d'))$ and $(x_j(\tilde{\theta}_i, R_j, \hat{\theta}_{-i,j}), y(\tilde{\theta}_i, R_j, \hat{\theta}_{-i,j})) = (x_j(c), y(c))$ where $(\tilde{\theta}_i, R_j, \hat{\theta}_{-i,j})$ is a preference profile corresponding to $(x_j(c), y(c))$. By construction, $\theta_i'' < \tilde{\theta}_i < \theta_i^*$. Since $y(c) < y(d')$, we have a contradiction to Lemma 2.

CLAIM 2: $\lim_{\theta_i\to\infty} x_i(\theta_i, \hat{\theta}_{-i}) \geq \omega_i$.
IRED for agent i requires

$$y(\theta_i, \hat{\theta}_{-i}) \geq \theta_i\left[\sqrt{\omega_i} - \sqrt{x_i(\theta_i, \hat{\theta}_{-i})}\right]. \tag{2.10}$$

Since $x_i(\theta_i, \hat{\theta}_{-i})$ is bounded and monotone in θ_i, the limit exists. If the limit is strictly less than ω_i ,then the RHS of Equation 2.10 increases without bound as $\theta_i \to \infty$. On the other hand, the LHS is bounded above by ω. We therefore have a contradiction.

CLAIM 3: $\lim_{\theta_i\to\infty} y(\theta_i, \hat{\theta}_{-i}) = 0$.
Suppose the claim is false i.e. $\lim_{\theta_i\to\infty} y(\theta_i, \hat{\theta}_{-i}) = \alpha > 0$. We proceed in several steps.

Step 1: There exists $j \neq i$ such that $\lim_{\theta_i\to\infty} x_j(\theta_i, \hat{\theta}_{-i}) < \omega_j$.

By Claim 1, $x_j(\theta_i, \hat{\theta}_{-i})$ is monotone. Therefore $\lim_{\theta_i\to\infty} \sum_{j\neq i} x_j(\theta_i, \hat{\theta}_{-i})$ exists. If the claim is false, then $\lim_{\theta_i\to\infty} \sum_{j\neq i} x_j(\theta_i, \hat{\theta}_{-i}) \geq \sum_{j\neq i} \omega_j$. However taking limits in the equation for feasibility and using Claim 2, we have

$$\lim_{\theta_i\to\infty} \sum_{j\neq i} x_j(\theta_i, \hat{\theta}_{-i}) \leq \sum_{j\neq i} \omega_j - c(\alpha).$$

By assumption, $c(\alpha) > 0$ which leads to a contradiction.

In the rest of the proof of Claim 3, j is chosen so that $\lim_{\theta_i\to\infty} x_j(\theta_i, \hat{\theta}_{-i}) < \omega_j$ holds. Let $\lim_{\theta_i\to\infty} x_j(\theta_i, \hat{\theta}_{-i}) = \beta < \omega_j$.

As before, let $\theta_i \mapsto (x_j(\theta_i, \hat{\theta}_{-i}), y(\theta_i, \hat{\theta}_{-i}))$ for $\theta_i > 0$ define a curve in (x_j, y) space. This curve is referred to as Curve A. By Claim 1, Curve A is downward sloping and is not degenerate. Note that for any two points a and b on Curve A such that $x_j(a) \neq x_j(b)$ and

$y(a) > y(b)$, then $\theta_i^b > \theta_i^a$ where $(x_j(b), y(b)) = (x_j(\theta_i^b, \hat{\theta}_{-i}), y(\theta_i^b, \hat{\theta}_{-i}))$ and $(x_j(a), y(a)) = (x_j(\theta_i^a, \hat{\theta}_{-i}), y(\theta_i^a, \hat{\theta}_{-i}))$.

Note that (β, α) is a point on the curve.

Fix θ_i^n. The curve $B(\theta_i^n)$ is defined by $\theta_j \mapsto (x_j(\theta_i^n, \theta_j, \hat{\theta}_{-i}), y(\theta_i^n, \theta_j, \hat{\theta}_{-i}))$ for $\theta_j > 0$. SP for agent j implies that the curve is downward sloping in (x_j, y) space.

Step 2: There exists θ_i^n such that $B(\theta_i^n)$ does not coincide with Curve A till (β, α). [11]

Suppose this is false i.e. $B(\theta_i^q)$ coincides with Curve A till (β, α) for all q.

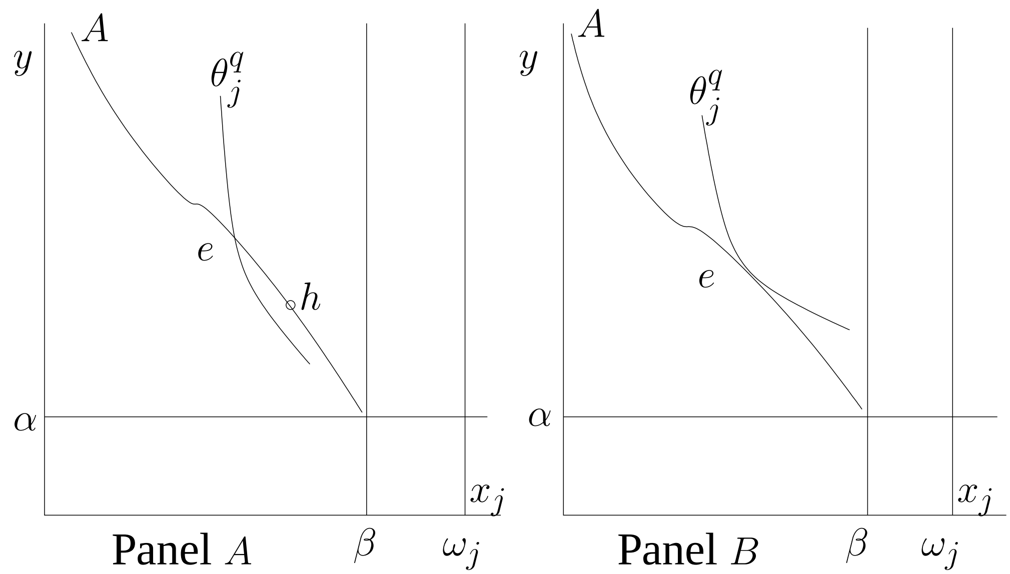

Figure 2.5:

We have assumed at the outset that $x_i(\hat{\theta}_i, \hat{\theta}_{-i}) > 0$. The allocation for agent j, $(x_j(\hat{\theta}_i, \hat{\theta}_{-i}), y(\hat{\theta}_i, \hat{\theta}_{-i}))$ lies on Curve A. Since A is not degenerate, we can find a point e on A, close to this point where the curve is differentiable (follows from the monotonicity of $x_j(\theta_i, \hat{\theta}_{-i})$ and $y(\theta_i, \hat{\theta}_{-i})$ in θ_i). Thus $x_i(e) > 0$.

By our hypothesis, it follows that for all θ_i^q, there exists θ_j^q such that the allocation for player j at e equals $(x_j(e), y(e)) = (x_j(\theta_i^q, \theta_j^q, \hat{\theta}_{-i,j}), y(\theta_i^q, \theta_j^q, \hat{\theta}_{-i,j}))$.

We argue that θ_j^q does not depend on θ_i^q. Pick an arbitrary θ_i^q. We claim that curve A must be tangent to indifference curve of preference θ_j^q. Suppose not. In particular, consider the indifference curve in Figure 2.5 (Panel A) and the allocation h for agent j. By hypothesis, h lies on $B(\theta_i^q)$ i.e. there exists $\tilde{\theta}_j$ such that $(x_j(h), y(h)) = (x_j(\theta_i^q, \tilde{\theta}_j, \hat{\theta}_{-i,j}), y(\theta_i^q, \tilde{\theta}_j, \hat{\theta}_{-i,j}))$. But then j manipulates at θ_j^q via $\tilde{\theta}_j$. We can conclude therefore that all the indifference curves corresponding to all the different θ_j^q have a common tangent at e. Thus for θ_j^q and $\theta_j^{q'}$, we have $\frac{\theta_j^q}{2\sqrt{x_j(e)}} = \frac{\theta_j^{q'}}{2\sqrt{x_j(e)}}$ which implies that $\theta_j^q = \theta_j^{q'}$. We refer to this common θ_j^q as $\bar{\theta}_j$.

[11] Note that if Curve A is vertical, then our claim is trivially true.

Now since e lies on $B(\theta_i^q)$ for all q, $y(\theta_i^q, \bar{\theta}_j, \hat{\theta}_{-i,j}) = y(e)$. Therefore SP for agent i implies that $x_i(\theta_i^q, \bar{\theta}_j, \hat{\theta}_{-i,j}) = x_i(e)$ (which is strictly positive).

By PERD

$$\frac{2\sqrt{x_i(e)}}{\theta_i^q} + \sum_{k \neq i} \alpha(\theta_i^q, \bar{\theta}_j, \hat{\theta}_{i,j}) \frac{2\sqrt{x_k(\theta_i^q, \bar{\theta}_j, \hat{\theta}_{-i,j})}}{\theta_k} = c'(y(e)) \ \forall \ q. \tag{2.11}$$

The LHS of Equation 2.11 becomes unboundedly large as $\theta_i^q \to 0$. On the other hand, the RHS is finite which leads to a contradiction. This establishes Step 2.

Pick (x_j', y') such that $\beta < x_j' < \omega_j$ and $0 < \alpha < y'$. Let $\theta_j^* = \frac{y'}{(\sqrt{\omega_j} - \sqrt{x_j'})}$. The indifference curve for ordering θ_j^* is shown in Figure 2.6. Let θ_i^n be the preference ordering of agent i guaranteed by Step 2 i.e. $B(\theta_i^n)$ does not coincide with Curve A till (β, α). IRED for agent j implies that $B(\theta_i^n)$ must enter in the shaded region in Figure 2.6. There are two cases to consider.

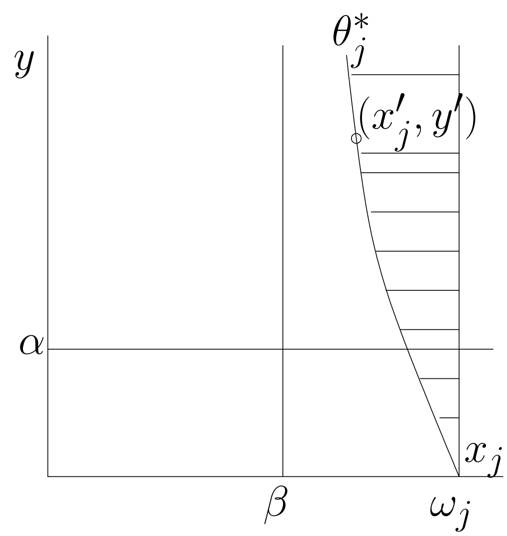

Figure 2.6: Existence of θ_j^*

Case (i): Some part of the Curve $B(\theta_i^n)$ lies below Curve A (Figure 2.7).
Case (ii): Curve $B(\theta_i^n)$ lies completely above Curve A (Figure 2.8).

We will show that each case leads to a contradiction.

Case (i): In Figure 2.7, pick points e and h on Curves $B(\theta_i^n)$ and A respectively such that $x_j(e) < x_j(h)$ and $y(e) < y(h)$ Let e and h correspond to the profiles $(\theta_i^n, \theta_j^e, \hat{\theta}_{-i,j})$ and $(\theta_i^h, \hat{\theta}_j, \hat{\theta}_{-i,j})$. By assumption and our earlier observations, it follows that $\theta_i^h > \theta_i^n$. Since $x_j(e) < x_j(h)$ and $y(e) < y(h)$, Lemma 3 there exists a common concavification R_j of θ_j^e at e and $\hat{\theta}_j$ at h (see Figure 2.7) [12]. By SP of agent j,

[12] Note that in this case, condition (iv) of Lemma 3 is satisfied trivially because $L(e, h)$ is upward sloping.

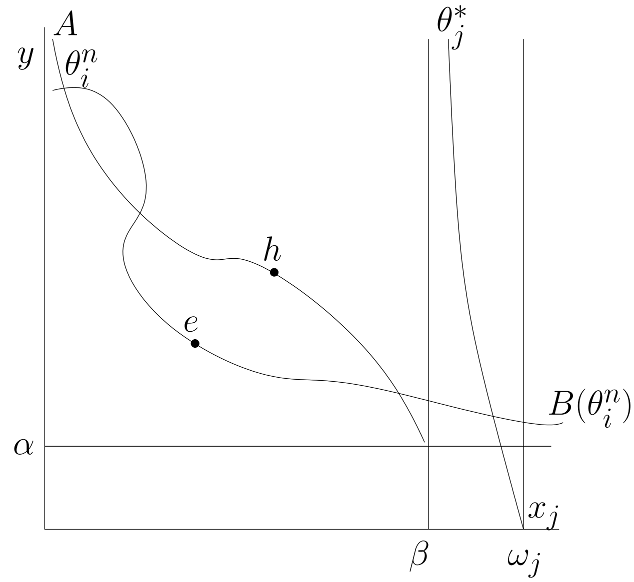

Figure 2.7: Case (i)

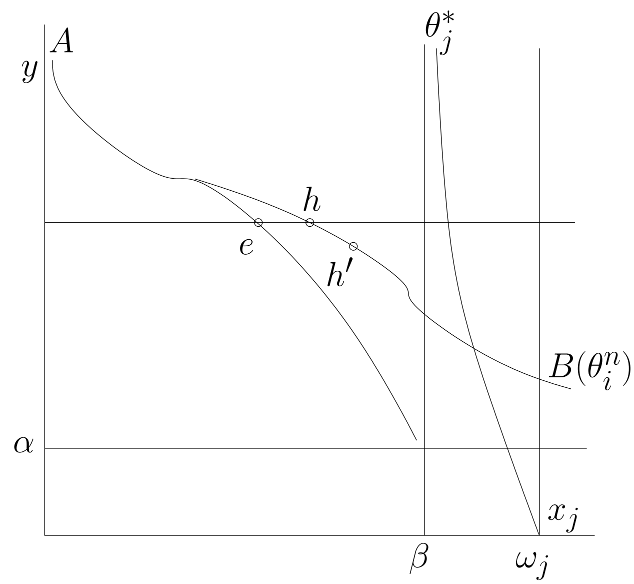

Figure 2.8: Case (ii)

1. $(x_j(\theta_i^h, \hat{\theta}_j, \hat{\theta}_{-i,j}), y(\theta_i^h, \hat{\theta}_j, \hat{\theta}_{-i,j})) = (x_j(\theta_i^h, R_j, \hat{\theta}_{-i,j}), y(\theta_i^h, R_j, \hat{\theta}_{-i,j}))$.

2. $(x_j(\theta_i^n, \theta_j^e, \hat{\theta}_{-i,j}), y(\theta_i^n, \theta_j^e, \hat{\theta}_{-i,j})) = (x_j(\theta_i^n, R_j, \hat{\theta}_{-i,j}), y(\theta_i^n, R_j, \hat{\theta}_{-i,j}))$.

Since $\theta_i^h > \theta_i^n$, we have a contradiction to non-increasingness of y in θ_i (Lemma 2).

Case (ii): This case proceeds in a manner very similar to Claim 1. Pick points e and h on Curves A and $B(\theta_i^n)$ such that $y(e) = y(h)$. Let h correspond to the profile $(\theta_i^n, \theta_j^{'}, \theta_{-i,j})$ and we denote $x_j(\theta_i^n, \theta_j^{'}, \theta_{-i,j}) = x_j(h)$. See Figure 2.8.

Consider h' suitably close to h on Curve $B(\theta_i^n)$ such that $y(h') < y(h)$. Arguing as we did in Claim 1, we can find $\theta_j^{h'}$ such that $(x_j(h'), y(h'))$ is strictly preferred to $(x_j(e), y(e))$ under both $\theta_j^{h'}$ and $\hat{\theta}_j$. We also need to ensure that $IC(\hat{\theta}_j, h')$ cuts $L(e, h')$ from above at point h'. Arguing as we did in Claim 1, such a point h' can be chosen. We can construct a strictly decreasing sequence of θ_j^k converging to $\theta_j^{'}$ and the corresponding points h^k moving closer to h as k increases. Thus, as θ_j^k approaches $\theta_j^{'}$, the slope (absolute value) of $L(e, h^k)$ tends to zero. Also, the limit of the slope of $IC(\theta_j^{'}, h^k)$ at point h^k is $\frac{\theta_j^{'}}{2\sqrt{x_j(h)}}$ which is strictly positive. Thus there exists K and h^K where condition (iv) of Lemma 3 will be satisfied. We choose $h' = h^K$. We can therefore find a common concavification R_j of $\theta_j^{h'}$ at h' and $\hat{\theta}_j$ at e and obtain a contradiction to the non-increasingness of y in θ_i.

These arguments establish Claim 3.

We now complete the proof of the theorem.

Choose $\tilde{\theta}_{-i}$ such that

$$\sum_{k \neq i} \frac{2\sqrt{\omega_k}}{\tilde{\theta}_k} > c'(0) \geq 0.$$

Consider a sequence of profiles $(\theta_i, \tilde{\theta}_{-i})$, $\theta_i \to \infty$. Since our initial choice of $\hat{\theta}_{-i}$ was arbitrary, we have

$$\lim_{\theta_i \to \infty} y(\theta_i, \tilde{\theta}_{-i}) = 0. \tag{2.12}$$

We have already shown that for θ_i large enough, $x_i(\theta_i, \tilde{\theta}_{-i}) > 0$. By IRED for agent $k \neq i$ and Equation 2.12, we can conclude that

$$\lim_{\theta_i \to \infty} x_k(\theta_i, \tilde{\theta}_{-i}) \geq \omega_k > 0 \quad \text{for all } k \neq i. \tag{2.13}$$

Hence, $x_k(\theta_i, \tilde{\theta}_{-i}) > 0$ for θ_i large enough. For these profiles, PERD requires

$$\frac{2\sqrt{x_i(\theta_i, \tilde{\theta}_{-i})}}{\theta_i} + \sum_{k \neq i} \frac{2\sqrt{x_k(\theta_i, \tilde{\theta}_{-i})}}{\tilde{\theta}_k} \leq c'(y(\theta_i, \tilde{\theta}_{-i})). \tag{2.14}$$

Taking limits as $\theta_i \to \infty$, we have

$$\lim_{\theta_i \to \infty} \sum_{k \neq i} \frac{2\sqrt{x_k(\theta_i, \tilde{\theta}_{-i})}}{\tilde{\theta}_k} \geq \sum_{k \neq i} \frac{2\sqrt{\omega_k}}{\tilde{\theta}_k} > c'(0).$$

However this contradicts our choice of $\tilde{\theta}_{-i}$ and completes the proof of the theorem.

■

2.5 Discussion

We would like to draw attention to certain features of the domain of preferences in our theorem. A SCF in our definition, is defined on the *full* domain of classical preferences. Although the PERD, IRED and RDC conditions are required to hold only on the (much smaller) restricted domain $\mathcal{D}^n$, strategy-proofness is required to hold over the full domain. However, as the proof of the theorem demonstrates, a much smaller domain suffices for our argument. The Appendix provides details for the additional preferences required. A natural question is whether our result is valid for the domain $\mathcal{D}^n$, i.e. whether strategy-proofness applied only to $\mathcal{D}^n$ is sufficient for the result. Unfortunately, we are unable to settle this issue presently, though we conjecture that it is true.

We make brief remarks regarding the necessity of our axioms. The *dictatorial* SCF chooses an allocation that maximizes the preference of a particular agent subject to production and feasibility constraints. This SCF satisfies PERD, strategy-proofness and RDC but violates IRED.

Serizawa (1996) proposes a SCF called the *convex cost sharing rule* that satisfies strategy-proofness, IRED and RDC but violates PERD.

The set of Pareto-efficient and individually rational allocations at any preference profile in the domain $\mathcal{D}^n$, is non-empty and compact. Any continuous selection from the Pareto-efficient and individually rational correspondence will satisfy all requirements of our theorem except strategy-proofness.

Finally, we make some remarks about our notion of individual rationality. Serizawa (1996) considers a richer model where each agent i has access to a private public good production technology with an associated cost function $c_i(.)$. There is also a jointly available production technology with a cost function $c(.)$. An agent has the option of not offering any part of his endowment for the joint production of the public good (using the cost function $c(.)$) and instead producing the public good with her own technology i.e. using $c_i(.)$. In our model, agents cannot produce any public good on their own. This implies that the reservation utility level of an agent in the Serizawa model is weakly higher than in our model i.e. his individual rationality constraints are more severe. Our impossiblity result will therefore continue to hold if our individual rationality constraint is substituted by Serizawa's individual rationality constraint. There is however an important caveat. We are assuming that the existence of agent specific production technologies does not affect the Pareto efficiency condition. For instance, Pareto efficiency could involve the splitting of the aggregate public good production into smaller units produced by agents using their individual technologies. This might happen even in the case when the $c_i(.)$ functions are the same as the $c(.)$ function, if the latter is strictly convex.

2.6 Appendix

We provide a formal proof of Lemma 3.

Proof: Let $(x_i^{'}, y^{'}), (x_i^{''}, y^{''}) \in \Re_+^2$ and $\theta_i, \theta_i^{'}$ be specified in accordance with the statement of Lemma 3. Assume without loss of generality, $x_i^{'} < x_i^{''}$ and $y^{'} > y^{''}$. We note that $\theta_i < \theta_i^{'}$, in all the sections where Lemma 3 is used.

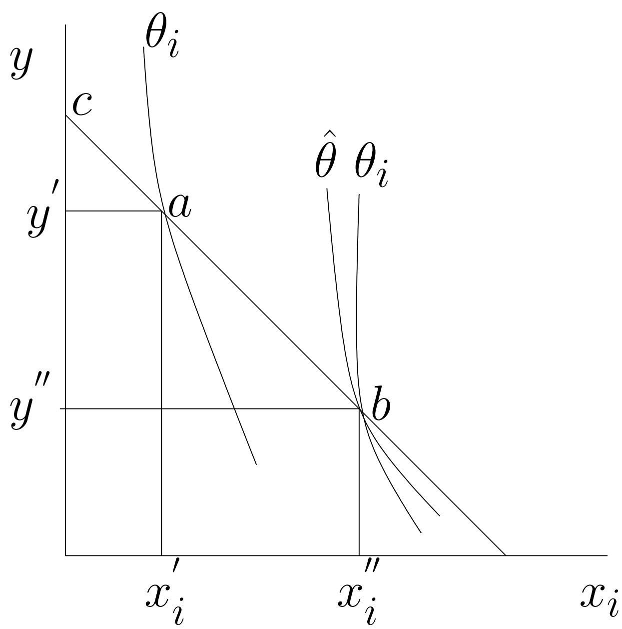

Figure 2.9: Construction of a common concavified preference

In Figure 2.9, the points $(x_i^{'}, y^{'})$ and $(x_i^{''}, y^{''})$ are denoted by a and b respectively. Let $L(a, b)$ denote the line joining a and b. The point of intersection of $L(a, b)$ and the y axis is denoted by c. In view of our hypothesis, $IC(\theta_i, a)$ is not tangential to $L(a, b)$ at a, i.e it must intersect $L(a, b)$ at a.

Choose $\tilde{\theta}$ such that $\tilde{\theta} > \theta_i^{'}$. A direct computation of slopes at a reveals that $IC(\tilde{\theta}, a)$ cuts $IC(\theta_i, a)$ from above at a. Also, $IC(\tilde{\theta}, b)$ cuts $IC(\theta_i^{'}, b)$ from above at b [13].

We denote the absolute value of the slope of $L(a, b)$ by $\bar{c}$. In view of our hypothesis, specifically condition (iv) we have,

$$\frac{\theta_i}{2\sqrt{x_i^{''}}} > \bar{c} \implies \theta_i > 2\bar{c}\sqrt{x_i^{''}}$$

Thus we can choose $\hat{\theta}$ such that $\theta_i > \hat{\theta} > 2\bar{c}\sqrt{x_i^{''}}$. It follows from our construction that $IC(\hat{\theta}, b)$ cuts $L(a, b)$ from above at b.

[13]The domain $\mathcal{D}$ is a single-crossing domain. These properties hold generally for such domains - for details see Goswami (2013).

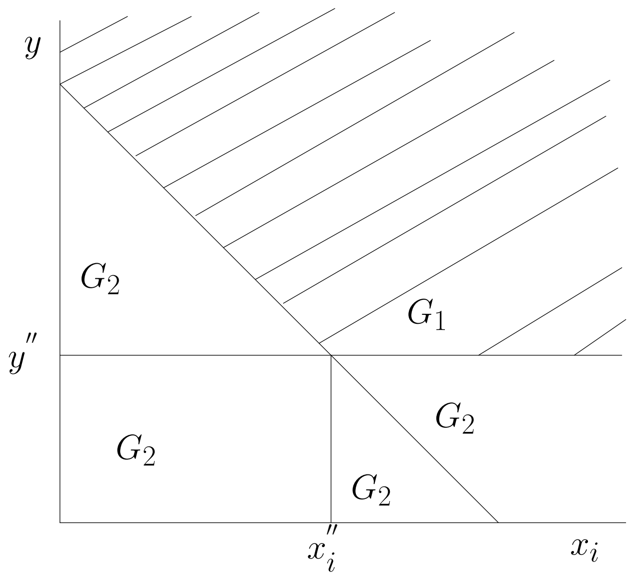

Figure 2.10: Construction of a common concavified preference

The parameters $\tilde{\theta}$ and $\hat{\theta}$ chosen above can be used to construct a preference R_i which is a concavification of θ_i at (x_i', y') and θ_i' at (x_i'', y''). The regions G_1 and G_2 are indicated in Figure 2.10. An additional parameter t is computed as follows.

Pick an arbitrary $(x_i, y) \in G_1$. Let $h = (x_i(h), y(h))$ be such that

(i) $\tilde{\theta}\sqrt{x_i} + y = \tilde{\theta}\sqrt{x_i(h)} + y(h)$ and

(ii) $[(x_i(h) < x_i'', y(h) > y'')$ and$(x_i(h), y(h))$ lies on $L(c, b)]$ or $[x_i(h) \geq x_i'', y(h) = y'']$.

Let $t = \frac{x_i(h)}{x_i'}$. The preference ordering R_i is defined below.

$$R_i(x_i, y) = \begin{cases} \tilde{\theta}\sqrt{x_i} + y - \sqrt{t x_i'}(\tilde{\theta} - \hat{\theta}), & \text{if } (x_i, y) \in G_1 \\ \hat{\theta}\sqrt{x_i} + y, & \text{if } (x_i, y) \in G_2 \end{cases}$$

We omit the verification of the following facts.

(i) R_i is a classical preference.

(ii) It concavifies θ_i at (x_i', y') and θ_i' at (x_i'', y'').

Requirement (iv) in the statement of Lemma 3 ensures that the indifference curves for R_i do not intersect each other.

■

Chapter 3

Selecting Winners with Partially Honest Jurors

3.1 Introduction

The theory of mechanism design investigates the goals that a planner can achieve when these goals depend on private information held by various agents. The planner designs a mechanism that elicits the private information held by various agents. The theory is based on the assumption that agents act purely to further their self-interest - an assumption that is common to most of the literature in economics. However, there is both empirical and experimental evidence that considerations other than self-interest influence individual behaviour. Several recent papers have considered departures from the standard implementation framework.[1] In this paper, we investigate the implication of one such behaviourial assumption - that of partial honesty in the model of juror decisions developed in Amorós (2010) and Amorós (2013). Our main result is that this assumption expands the scope of implementation when there are three or more jurors, but has no effect when there are exactly two jurors.

We briefly describe the juror model. A set of $n \geq 3$ contestants are involved in a competition. A jury must choose a set of $w < n$ contestants who will win the competition. All jurors know who the w best contestants are. We refer to this set as the set of deserving winners. Each juror, may be biased in favour of (or against) some contestants. Amorós (2010) shows the following necessary condition for implementation. Fix a pair of contestants: there must exist a juror who is fair over this pair and the identity of this juror must be known to the planner. We say that a juror is fair over a pair of contestants, if while comparing two sets of winners, which differ exactly in this pair, the juror strictly prefers the set which contains the contestant which belongs to the deserving set of winners. Amorós (2010) establishes that this condition is also sufficient for implementation. We note that the condition that

[1] See also Kartik and Tercieux (2012), Dutta and Sen (2012), Matsushima (2008), Lombardi (2010) and Lombardi and Yoshihara (2011).

the identity of the fair juror is known to the planner , is a strong one. We show that in the presence of partially honest agents, this condition is no longer necessary.

Following Dutta and Sen (2012), we assume that there are some jurors who have an intrinsic preference for honesty. These jurors have preferences not only on outcomes but also on the *messages* that they are required to send to the planner. Suppose the mechanism used by the planner requires each juror to announce a set of winners. Then a juror is partially honest , if she strictly prefers to be truthful when this announcement does not change the outcome (given the messages announced by others). Thus the partially honest jurors preferences are lexicographic since the preference for honesty is operational only when the juror is indifferent on the outcome dimension.

An important feature of our definition of partial honesty is that it applies only to contestant pairs over which he is fair. Suppose an agent is fair and partially honest over the pair (a, b). We illustrate this with an example. Suppose $w = 2$. Fairness implies that the juror prefers the set $\{a, c\}$ to $\{b, c\}$ for all c whenever a is in the deserving set and b is not. Partial honesty on the other hand implies the following: suppose a is the deserving set and b is not. Suppose a truthful message (one which involves announcing the deserving set of winners, that includes a) and a non-truthful message both lead to the outcome bc in a mechanism. Then the juror strictly prefers sending the truthful message.

Our main results are the following. We know from Amorós (2010)) that if the identity of a juror who is fair over an arbitrary contestant pair is known to the planner then implementation is possible. Suppose this condition fails. We show that for every contestant pair, there must exist *two* jurors who are fair over that pair. Note that it is not necessary for the planner to know the identity of these two fair jurors. If one of these jurors is also partially honest then implementation is possible. We provide examples to show that implementation is possible even when the Amorós (2010)) condition fails. We show that the enhanced possibility result does not extend to the two person case.

We believe that it is appropriate to restrict partial honesty to contestant pairs over which a juror is fair. This is a natural assumption because the juror is fair over a pair presumably because she “cares” about the pair. We think of partial honesty therefore as an independent strengthening of the fairness condition. In Section 3.6, we show by means of an example that fairness over pairs cannot be substituted by extending partial honesty to pairs over which the juror is not fair. In the example, all jurors are partially honest over all pairs. However they are not fair over a specific contestant pair and implementation fails.

3.1.1 Related Literature

Amorós (2013) proposes a generalization of Amorós (2010). It reduces to the model in Amorós (2010) in a special case. Amorós (2009) analyzes a model where the jurors have to choose a full ranking of the contestants instead of selecting one winner. This paper provides necessary and sufficient conditions on the jury for the Nash implementability of the rule that always selects the socially optimal ranking. Amorós et al. (2002) study the same model and analyze implementation when each juror has one friend and is impartial with respect to the rest of contestants. Amorós (2011) studies a similar problem for subgame perfect implementation. Amorós (2014) studies conditions on the configuration of the jury when attention is restricted to simple and "natural" mechanisms.

Several papers examine the implementation of socially optimal rules under the assumption that some of the agents are *responsible*. Dogan (2013) studies a model , where a set of tasks is to be allocated among a set of agents whose preferences over allocations may or may not be responsive to the optimal allocation. The notion of an agent being responsible in this paper is similar to that of a juror being fair in our model. This paper shows that the optimal allocation can be implemented in Nash equilibrium if there are at least three responsible agents.

In the next section, we describe the model and notation. Section 3.3 introduces the concept of partially honest jurors. Sections 3.4 and 3.5 present results pertaining to the many-person and the two-person implementation problems respectively. Section 3.6 discusses the relationship between partial honesty and fairness. Section 3.7 concludes.

3.2 The Model

We closely follow the model of Amorós (2010) and Amorós (2013).

There is a set $N = \{a, b, c, ...\}$, $|N| \geq 3$, of contestants in a competition. A group of jurors $J = \{1, 2, \ldots, |J|\ \}$ must choose a subset W of winners, where $|W|\ = w$ and $0 < w < n$. Each juror i has a preference ordering R_i over the set of all possible winners, which are all subsets of N of size w. Let 2^N_w denote the set of all possible subsets of N of size w and let $\mathcal{R}$ denote the set of all orderings over 2^N_w.

It is assumed that there is a set of deserving winners $W^* \in 2^N_w$, which is commonly known to all jurors. Each juror i's ordering depends on W^*, i.e. juror i has a preference function $R_i : 2^N_w \to \mathcal{R}$. The ordering $R_i(W^*)$ reflects biases that juror i has over contestants. For instance, $R_i(W^*)$ may not have W^* as a maximal element because i is biased in favour of a candidate not included in W^* or biased against a candidate included in W^*. The goal of the mechanism designer is to implement W^* irrespective of the preferences of jurors.

We begin by describing admissible preferences for jurors. In what follows, $P_i(W^*)$ denotes the asymmetric component of $R_i(W^*)$.

Let N^2 be the collection of all possible pairs of contestants. Fix $i \in J$. Let $F_i \subset N^2$, i.e. F_i is a collection of pairs of contestants $(a, b), (c, d)$ etc.

Definition 7 *Fix $a, b \in N$. The sets $W, W' \in 2^N_w$ are said to be (a, b) variant if*

1. *$a \in W$.*

2. *$b \in W'$.*

3. $W \setminus \{a\} = W' \setminus \{b\}$.

Definition 8 *The preference function* $R_i : 2^N_w \to \mathcal{R}$ *is admissible at* F_i*, if for each pair* $(a, b) \in F_i$*, each* $W^* \in 2^N_w$*, and each* $W, W' \in 2^N_w$ *with*

(i) W *and* W' *are* (a, b) *variant*

(ii) $a \in W^*$

(iii) $b \notin W^*$,

we have $W \ P_i(W^*) \ W'$.

If $(a, b) \in F_i$ and R_i is admissible at F_i, we say that juror i *treats* (a, b) *fairly*. We shall refer to F_i as the fair set for juror i. Suppose W^* is the deserving set of winners and a belongs to W^* but b does not. Then for any $W, W' \in 2^N_w$ such that W and W' differ only in a and b, juror i strictly prefers W over W'. Note that several orderings may be admissible with respect to F_i. The set of preference orderings admissible with F_i is denoted by $\mathcal{R}(F_i)$. We illustrate this with an example.

Example 1 Let $N = \{a, b, c, d\}$, $w = 3$ and $i \in J$. Then $2^N_w = \{abc, abd, acd, bcd\}$. Suppose $F_i = \{(a, b)\}$. Any preference function that is admissible at F_i, $R_i \in \mathcal{R}(F_i)$ will satisfy the following restrictions:

1. When $W^* = acd$, the only restriction on $R_i(acd)$ is $acd \ P_i(acd) \ bcd$.

2. When $W^* = bcd$, the only restriction on $R_i(bcd)$ is $bcd \ P_i(bcd) \ acd$.

Note that no restrictions are imposed on $R_i(W^*)$ when $W^* \in \{abc, abd\}$, since both contestants in the pair (a, b) belong to W^*. Also when comparing abc with abd, the only two contestants whose winner status changes are c and d. Since $(c, d) \notin F_i$, both rankings, $abc \ R_i(W^*) \ abd$ and $abd \ R_i(W^*) \ abc$ are admissible for any W^*. When comparing abc and acd, the only two contestants whose winner status changes are b and d. Since $(b, d) \notin F_i$, both rankings $abc \ R_i(W^*) \ acd$ and $acd \ R_i(W^*) \ abc$ are admissible for any W^*. This is summarized in Table 3.1 below.

Let $F \equiv (F_i)_{i \in J}$ denote a profile of fair sets of the jurors. The planner is uncertain about the fair sets. This uncertainty is represented by the set Ω where $\Omega = \{F, F', ...\}$. The interpretation of the structure Ω is that the planner is unaware of the realization of the fair set in Ω.

A state of the world is a pair $(R, W^*) \in \mathcal{R}^{|J|} \times 2^N_w$ where $R \equiv (R_i)_{i \in J} \in \times_{i \in J} \mathcal{R}(F_i)$ for some $F \in \Omega$. Let $S(\Omega)$ be the set of the admissible states of the world, i.e.

$R_i(abc)$	$R_i(abd)$	$R_i(acd)$	$R_i(bcd)$
		.	.
		.	.
No restriction	No restriction	acd	bcd
		.	.
		.	.
		bcd	acd

Table 3.1: Example 1

$$S(\Omega) = \left\{(R, W^*) \in \mathcal{R}^{|J|} \times 2^N_w : \exists F \in \Omega \ s.t. \ R \equiv (R_i)_{i\in J} \in \times_{i\in J}\mathcal{R}(F_i)\right\}$$

The socially optimal rule is the function $\varphi : S(\Omega) \to 2^N_w$ where $\varphi(R, W^*) = W^*$ for each $(R, W^*) \in S(\Omega)$. The socially optimal rule selects the deserving winners for each admissible state.

A mechanism is a pair $\Gamma \equiv (M, g)$, where $M = \times_{i\in J} M_i$, M_i is the message space for juror i and $g : M \to 2^N_w$ is an outcome function. A message $m \in M$ is a Nash equilibrium of $\Gamma \equiv (M, g)$ at $(R, W^*) \in S(\Omega)$ if for each $i \in J$ and each $\hat{m}_i \in M_i$, $g(m)\ R_i(W^*)\ g(\hat{m}_i, m_{-i})$. Let $N(\Gamma, R, W^*) \subseteq M$ denote the set of Nash equilibria of Γ at (R, W^*). The mechanism must select the deserving winners in equilibrium in each state. We state this formally below.

Definition **9** *The mechanism $\Gamma \equiv (M, g)$ Nash-implements φ if for each $(R, W^*) \in S(\Omega)$*

(i) there exists $m \in N(\Gamma, R, W^*)$ such that $g(m) = W^*$ and

(ii) if $m \in M$ is such that $g(m) \neq W^*$, then $m \notin N(\Gamma, R, W^*)$.

If such a mechanism exists, then φ is Nash-implementable.

The remark below highlights an important difference between our model and that of Amorós (2010) and Amorós (2013).

REMARK: Amorós (2013) and Amorós (2010) define the fair set as a subset of the set of contestants. One implication of his formulation is that if i treats (a, b) and (b, c) fairly, then i treats (a, c) fairly. This is not true for our model as we allow for the possibility that i may not treat (a, c) fairly while treating (a, b) and (b, c) fairly. We believe that our assumption is more natural if one believes that treating contestants fairly is a matter of personal characteristics. Our assumption permits a larger class of admissible preferences. Our assumption also has a bearing on the two person case which we will comment on in Section 5.

3.3 Partially Honest Jurors

The classical literature on implementation assumes that individuals are fully strategic and care only about the realized outcomes. Some recent papers consider departures from the standard model by assuming that the agents are not fully rational. In particular some agents may have a preference for honesty.[2]

Dutta and Sen (2012) define a notion of honesty (called partial honesty) and consider the implications of this assumption for the standard implementation model. An individual is said to be *partially honest* if she strictly prefers to be truthful whenever a lie does not affect her material payoff. This restriction is weak since payoff considerations lexicographically dominates the desire to be truthful. The concept used in our paper links unbiasedness to honesty. It requires a partially honest individual to be honest only over winner sets that differ only over pairs of contestants which the juror treats fairly.

We focus on mechanisms in which one component of each individual's message space is the announcement of the set of winners. Consider a mechanism Γ where $M_i = 2^N_w \times S_i$ for each $i \in J$ and S_i denotes the other components of the message space. Following Dutta and Sen (2012), an individual's ordering over 2^N_w can be extended to an ordering $\succsim_i^{W^*}$ over the message space M. The asymmetric component of $\succsim_i^{W^*}$ will be denoted by $\succ_i^{W^*}$.

Definition **10** *A juror i is F_i-partially honest if for each pair $(a, b) \in F_i$, each $W^* \in 2^N_w$ and each $W, W' \in 2^N_w$ satisfying*

(i) $a \in W^*$, $b \notin W^*$.

(ii) W *and* W' *are* (a, b) *variant.*

(iii) $(m_i, m_{-i}), (m_i', m_{-i}) \in M$ *with*

- *(a)* $m_i \in \{W\} \times S_i$
- *(b)* $m_i' \in \{W'\} \times S_i$
- *(c)* $g(m_i, m_{-i}) = g(m_i', m_{-i}) = W'$

we have $(m_i, m_{-i}) \succ_i^{W^*} (m_i', m_{-i})$.
In all other cases, $(m_i, m_{-i}) \sim_i^{W^*} (m_i', m_{-i})$ *if* $g(m_i, m_{-i}) = g(m_i', m_{-i})$.

The juror (agent) may be honest only over the pairs of contestants that he treats fairly. Thus he is honest only when comparing sets of winners that differ exactly in the pair that he treats fairly and when one of the contestants in the pair is in W^* (while the other is not). Suppose juror i treats the pair (a, b) fairly. Let $a \in W^*$ and $b \notin W^*$. Then for all sets W, W' that are (a, b) variant, juror i's preference is $WP_i(W^*)W'$. Consider messages $m, m' \in M$ of juror i. Suppose m and m' involve the announcement of W and W' respectively and

[2] See Kartik and Tercieux (2012).

the outcome of both messages is W'. If juror i is partially honest, he strictly prefers the message m over m' even though both result in the outcome W'. Note that F_i-partial honesty is weaker than partial honesty.

3.4 Many Person Implementation

Amorós (2013) provides a necessary and sufficient condition for Nash implementation. The condition is as follows: For each $(a, b) \in N^2$, there exists a juror i who is known to the planner to be fair over that pair. Formally

Definition **11** *The structure Ω satisfies Condition A if for each $(a, b) \in N^2$, there exists $i \in J$ such that $(a, b) \in F_i$ for all $F \in \Omega$.*

Since $(a, b) \in F_i$ for all $F \in \Omega$, it follows that the planner *knows the identity* of the juror who is fair over a given pair. The goal of this section is to show that the condition for implementability can be significantly weakened if some jurors are partially honest.

Definition **12** *The structure Ω satisfies Condition B if for each $F \in \Omega$ and $(a, b) \in N^2$,*

(i) There exists $i, j \in J$ such that $(a, b) \in F_i$ and $(a, b) \in F_j$.

(ii) There exists $k \in \{i, j\}$ such that k is F_k-partially honest.

The structure Ω satisfies Condition B if for any given pair there exist two jurors who are fair with respect to it. In addition, one of the fair jurors is also partially honest. In contrast to Condition A, the identities of the fair/partially honest jurors *need not* be known to the planner. Our first result shows that the Condition B is sufficient for implementation.

Theorem **2** *Suppose that there are at least three jurors and Ω satisfies Condition B. Then φ is Nash implementable.*

Proof: Let $\Gamma \equiv (M, g)$ be a mechanism where for each $i \in J$, $M_i = 2^N_w \times \{1, 2, ..., |J|\}$. The outcome function is specified by the following rules:

Rule 1: If for each $i \in J$, $(W_i, z_i) = (W, z)$, then $g(m) = W$.

Rule 2: If there is $j \in J$ such that

1. $(W_i, z_i) = (W, z)$ for each $i \neq j$.

2. $(W_j, z_j) \neq (W, z)$.

then $g(m) = W$.

Rule 3: In all the other cases, $g(m) = W_j$ for $j \in J$ such that $j = (\sum_{i \in J} z_i)(mod\,|J|)$.

We will show that the mechanism Γ Nash-implements the socially optimal rule.

Claim 1: For each $(R, W^*) \in S(\Omega)$, there exists $m \in N(\Gamma, R, W^*)$ such that $g(m) = W^*$.

Let $(R, W^*) \in S(\Omega)$. Consider message $m = ((W_i, z_i))_{i \in J}$ where for each $i \in J$, $(W_i, z_i) = (W^*, n)$ for some integer n. Then Rule 1 of the mechanism is applicable and $g(m) = W^*$. We argue that $m \in N(\Gamma, R, W^*)$.

Fix $j \in J$. Consider a deviation by agent j to message $\hat{m}_j = (\hat{W}_j, \hat{z}_j)$. By Rule 2, $g(\hat{m}_j, m_{-j}) = W^*$, i.e. $m \in N(\Gamma, R, W^*)$.

Claim 2: For each $(R, W^*) \in S(\Omega)$ and $m \in M$ such that $g(m) \neq W^*$, we have $m \notin N(\Gamma, R, W^*)$.

Let $(R, W^*) \in S(\Omega)$. Fix an $F \in \Omega$ such that R is admissible with respect to F. Let $m \in M$ be such that $g(m) = W \neq W^*$. There are several cases to consider.

Case 1: Rule 1 applies to m. Then $m_i = (W, z)$ for each $i \in J$.

Since $W \neq W^*$, there exists $(a, b) \in N^2$ such that (i) $a \in W^*$, $a \notin W$ and (ii) $b \notin W^*$, $b \in W$. For an arbitrary $F \in \Omega$ and the pair (a, b), we know from Condition B: there exist $i, j \in J$ such that $(a, b) \in F_i$ and $(a, b) \in F_j$. We assume without loss of generality that juror i is F_i partially honest.

Let $\hat{W} \in 2^N_w$ be such that $(\hat{W}, W)$ are (a, b) variant. Consider a unilateral deviation by agent i from the message (m_i, m_{-i}) to $\hat{m}_i = (\hat{W}, z_i)$. Rule 2 is applicable for the message $(\hat{m}_i, m_{-i})$ and $g(\hat{m}_i, m_{-i}) = W$. Since $g(\hat{m}_i, m_{-i}) = g(m_i, m_{-i}) = W$ and i is F_i-partially honest, we have $(\hat{m}_i, m_{-i}) \succ_i^{W^*} (m_i, m_{-i})$, and so $m \notin N(\Gamma, R, W^*)$.

Case 2: Rule 2 applies to m. There exists $k \in J$ such that for $i \neq k$, $(W_i, z_i) = (W, z)$ and $(W_k, z_k) \neq (W, z)$. Since $W \neq W^*$, there exists $(a, b) \in N^2$ such that (i) $a \in W^*$, $a \notin W$ and (ii) $b \notin W^*$, $b \in W$. From Condition B, there exist $i, j \in J$ such that $(a, b) \in F_i$ and $(a, b) \in F_j$. Therefore there exist $h \neq k$ and $(a, b) \in F_h$. Let $\hat{W} \in 2^N_w$ be such that $(\hat{W}, W)$ are (a, b) variant. Consider a unilateral deviation by juror h to $\hat{m}_h = (\hat{W}, \hat{z}_h)$ where $h = (\hat{z}_h + \sum_{q \neq h} z_q)(mod\,|J|\ )$. Rule 3 is applicable for the message $(\hat{m}_h, m_{-h})$ and $g(\hat{m}_h, m_{-h}) = \hat{W}$. Since $(a, b) \in F_h$, we have $\hat{W} P_h(W^*) W$. Therefore $m \notin N(\Gamma, R, W^*)$.

Case 3: Rule 3 applies to m and $g(m) = W \neq W^*$.
Since $W \neq W^*$, there exists $a, b \in N$ such that (i) $a \in W^*$, $a \notin W$ and (ii) $b \notin W^*$, $b \in W$. From Condition B, we know that there exist $i, j \in J$ such that $(a, b) \in F_i$ and $(a, b) \in F_j$. Let $\hat{W} \in 2^N_w$ be such that $(\hat{W}, W)$ are (a, b) variant. Consider a unilateral deviation by juror

i to $\hat{m}_i = (\hat{W}, \hat{z}_i)$ where $i = (\hat{z}_i + \sum_{k \neq i} z_k)(mod\,|J|\ )$. Rule 3 is applicable for $(\hat{m}_i, m_{-i})$ and the outcome is $g(\hat{m}_i, m_{-i}) = \hat{W}$. Since $\hat{W}\ P_i(W^*)\ W$, we have $m \notin N(\Gamma, R, W^*)$.

Cases 1, 2 and 3 are exhaustive. Therefore Claim 2 is established and the proof is complete. ■

Conditions A and B are not comparable with respect to the fair sets of jurors for contestant pairs. While Condition A requires the identity of the fair juror over a pair to be known to the planner, Condition B does not. However Condition B requires the existence of two fair jurors for every contestant pair. They are equivalent only in the special case of exactly three jurors.

We show below that aspects of Condition B are necessary for implementation.

Theorem **3** *Suppose φ is Nash implementable in the presence of partially honest jurors. Fix $(a, b) \in N^2$. Then*

1. *For each $F \in \Omega$ and $(a, b) \in N^2$, there exists $i \in J$ such that $(a, b) \in F_i$.*
2. *Suppose Condition A does not hold. Then for each $F \in \Omega$, there exist $i, j \in J$ such that $(a, b) \in F_i$ and $(a, b) \in F_j$.*

Proof: We first prove Part 1 of the necessary condition.

Let φ be implementable by a mechanism $\Gamma = (M, g)$. Suppose there exists $F \in \Omega$ and $(a, b) \in N^2$ such that $(a, b) \notin F_i$ for all $i \in J$. Let $W^*, \hat{W}^* \in 2^N_w$ be (a, b) variant. Then there exists $R \equiv (R_i)_{i \in J} \in \times_{i \in J} \mathcal{R}(F_i)$ such that $R_i(W^*) = R_i(\hat{W}^*)$ for all i (see Table 1 for an illustration of such preferences). By (i) in the definition of Nash implementability, there exists $m \in N(\Gamma, R, W^*)$ with $g(m) = W^*$. The definition of F_i-partial honesty implies that there does not exist any juror i who is partially honest over (a, b). We claim that $m \in N(\Gamma, R, \hat{W}^*)$. To see this, consider a unilateral deviation $m_i^{'}$ by player i. If $g(m_i^{'}, m_{-i}) = W' \neq W^*$, then $W^* R_i(W^*) W'$. Since $R_i(W^*) = R_i(\hat{W}^*)$ and considerations of partial honesty do not apply (since $W' \neq W^*$), we have $m \in N(\Gamma, R, \hat{W}^*)$. Suppose $g(m_i^{'}, m_{-i}) = W^*$. Because $R_i(W^*) = R_i(\hat{W}^*)$ and considerations of partial honesty are not applicable, $m \in N(\Gamma, R, \hat{W}^*)$, which leads to a contradiction.

Suppose Part 2 of the necessary condition is violated. This implies that there exists $(a, b) \in N^2$ such that (i) there exist $F \in \Omega$ and $i \in J$ with $(a, b) \in F_i$ and $(a, b) \notin F_j$ for all $j \neq i$ and (ii) there exists $\hat{F} \in \Omega$ with $(a, b) \notin \hat{F}_i$ (since A does not hold).

Pick $W^*, \hat{W}^* \in 2^N_w$ that are (a, b) variant. We claim that there exist (i) $R \in \mathcal{R}(F)$ where R is illustrated in Table 3.2 below and (ii) $\hat{R} \in \mathcal{R}(\hat{F})$ where $\hat{R}$ is illustrated in Table 3.3 below.

Let $W^* = \{a, x_1, \ldots, x_K\}$ and $\hat{W}^* = \{b, x_1, \ldots, x_K\}$. Construct a function $u_i : N \to \mathbb{R}$ satisfying the following properties:

		R_i			R_j	
$W^* =$	.	W^*	.	.	W^*	.
	.	W^*	.	.	$\hat{W}^*$	.
Pref.	.	$\hat{W}^*$	.	.	.	.
	.	.	.	.	.	.

Table 3.2: $R \in \mathcal{R}(F)$

		$\hat{R}_i$	
$W^* =$	.	$\hat{W}^*$	.
	.	W^*	.
Pref.	.	.	.
	.	.	.

Table 3.3: $\hat{R} \in \mathcal{R}(\hat{F})$

1. $u_i(p) > u_i(q)$ for all $p \in W^*$ and $q \notin W^*$.

2. $u_i(p) > u(a)$ for all $p \in W^* \setminus \{a\}$.

3. $u_i(b) > u(q)$ for all $q \notin N \setminus W^* \cup \{b\}$.

Define $R_i(W^*)$ as follows: for any $W, W' \in 2^N_w$, $WR_i(W^*)W'$ iff $\sum_{p \in W} u_i(p) \geq \sum_{q \in W'} u_i(q)$. A suitable perturbation of u_i will make the ordering $R_i(W^*)$ anti-symmetric. Observe that these preferences are consistent with $(a, b) \in F_i$. Note that $R_i(W^*)$ is consistent with any $F_i \subseteq N^2$. Let $R_i(\hat{W}^*) = R_i(W^*)$. Note that this is consistent with any $\hat{F}_i$ with $(a, b) \in \hat{F}_i$. Hence $R_i(\hat{W}^*)$ is admissible with respect to $\hat{F}_i$.

Define $u_j : N \to \mathbb{R}$ by replacing a and b in u_i. Construct $R_j(W^*)$ from u_j in the same way as $R_i(W^*)$ was constructed from u_i. Observe that $\hat{W}^*$ is $R_j(W^*)$ maximal and admissible with respect to F_j.

By the definition of implementation, there exists $m(\hat{W}^*) \in M$ such that $g(m(\hat{W}^*)) = \hat{W}^*$ and $m(\hat{W}^*) \in N(\Gamma, \hat{R}, \hat{W}^*)$. Consider $(R, W^*) \in S(\Omega)$ where $R \in \mathcal{R}(F)$. Since $g(m(\hat{W}^*)) \neq W^*$, it must be the case that $m(\hat{W}^*) \notin N(\Gamma, R, W^*)$ by part (ii) of the definition of Nash implementability. Consider $j \neq i$. The definition of partial honesty implies that j is not partially honest over the pair (a, b). There does not exist any $m_j \in M$ such that a unilateral deviation to m_j is profitable for player j (since $\hat{W}^*$ is $R_j(W^*)$-maximal and considerations of partial honesty do not apply).

We claim that there exists $\tilde{m}_i \in M_i$ such that $g(\tilde{m}_i, m_{-i}(\hat{W}^*)) = W^*$. To see this, consider a unilateral deviation to $m_i^{'}$ by i. If $g(m_i^{'}, m_{-i}(\hat{W}^*)) = \hat{W}$ and $m_i^{'}$, $m_i(\hat{W}^*)$ are messages where W^* and $\hat{W}^*$ are announced respectively, then i benefits from truth telling and the unilateral deviation to $m_i^{'}$ is profitable. Thus $m(\hat{W}^*) \notin N(\Gamma, R, W^*)$. We now argue that $(m_i^{'}, m_{-i}(\hat{W}^*)) \notin N(\Gamma, R, W^*)$. The message profile $(m_i^{'}, m_{-i}(\hat{W}^*))$ is such that (i) $m_i^{'}$ is the message with the announcement W^* and (ii) $g(m_i^{'}, m_{-i}(\hat{W}^*)) = g(m(\hat{W}^*)) = \hat{W}^*$. Any message $m_i^{''}$ that keeps the outcome fixed at $\hat{W}$ and involves the announcement of W cannot be compared to $m_i^{'}$ using considerations of partial honesty. A message $m_i^{''}$ that results in $g(m_i^{''}, , m_{-i}(\hat{W}^*)) = W' \neq \hat{W}^*$ will be a profitable deviation for i if $W' = W^*$ (because $W^* P_i(W^*) \hat{W}^*$ and $\hat{W}^*$ is the second best alternative in $R_i(W^*)$). Thus there exists $\tilde{m}_i \in M_i$ such that $g(\tilde{m}_i, m_{-i}(\hat{W}^*)) = W^*$, which leads to a contradiction.

Since $g(\tilde{m}_i, m_{-i}(\hat{W}^*)) = W^*$ and $W^* \; \hat{P}(\hat{W}^*) \; \hat{W}^*$, we have $m(\hat{W}^*) \notin N(\Gamma, \hat{R}, \hat{W}^*)$. This contradicts the assumption that Γ implements φ.

■

3.4.1 Discussion

Our necessary condition can be interpreted in the following way. If Condition A holds, Amoros's result implies that φ is implementable. This result continues to hold if some of the jurors are partially honest - the same mechanism will continue to work. According to our result, implementation is possible *even if* Condition A fails. Theorem 3 shows that there must exist at least two jurors who are fair over every pair. If one of these jurors is also partially honest, Theorem 2 shows that φ is implementable. We conclude that the scope for implementation is enhanced significantly.

The example below demonstrates the existence of situations where implementation is possible even when Condition A fails.

Example 2 *Let $N = \{a, b, c\}$ and $J = \{1, 2, 3, 4\}$. The set $\Omega = \{F, \hat{F}\}$, where F and $\hat{F}$ are described in Table 3.4 and 3.5.*

F_1	F_2	F_3	F_4
$\{(a, b), (a, c)\}$	$\{(b, c), (a, b)\}$	$\{(a, c)\}$	$\{(b, c)\}$

Table 3.4: The collection F

In addition, all jurors $k \in \{1, 2, 3\}$) are assumed to be F_k-partially honest.

Fix an arbitrary pair of contestants. Observe that the planner does not know the identity of a juror who is fair over that pair. For instance $\{1, 2\}$ are fair over (a, b) in F (but not in $\hat{F}$) and $\{3, 4\}$ are fair over (a, b) in $\hat{F}$ (but not in F). Therefore the structure Ω violates Condition A. However Ω satisfies Condition B and hence is implementable.

$\hat{F}_1$	$\hat{F}_2$	$\hat{F}_3$	$\hat{F}_4$
$\{(b,c)\}$	$\{(a,c)\}$	$\{(a,b),(b,c)\}$	$\{(a,b),(a,c)\}$

Table 3.5: The collection $\hat{F}$

3.5 Two person Implementation

It is well known in implementation theory that the two person case has to be treated seperately from the three or more person case. The reason for this is that an additional incentive compatibility condition known as the *intersection property* becomes necessary for implementation. Details can be found in Dutta and Sen (1991), Moore and Repullo (1990) and Busetto and Codognato (2009).

Our main objective in this section is to show that the assumption of partial honesty does not enhance the scope of implementation. We proceed as follows. We first derive a necessary and sufficient condition for implementation in the standard model. Amorós (2013) does not treat the two person case for reasons that are discussed later. Then we show that this condition remains necessary even when both jurors are assumed to be partially honest.

Theorem 4 *Assume there are two jurors who are not partially honest. Then Condition A is necessary and sufficient for Nash implementation.*

Proof: The necessary and sufficient condition for implementation of φ when there are only two jurors is condition A, i.e. for each $(a,b) \in N^2$, there exists $i \in J$ such that $(a,b) \in F_i$ for all $F \in \Omega$.

The necessity of Condition A follows from the arguments in Amorós (2013) and Amorós (2010) for $|J| \geq 3$ case.[3] We therefore omit the proof of necessity of Condition A. The proof of sufficiency requires a new argument. We first introduce a definition that is important for the construction of the mechanism. In what follows the two jurors will be referred to as i and j.

Let $W, \hat{W} \in 2^N_w$ be such that $W \neq \hat{W}$. We say that $(W, \hat{W})$ satisfies *Property* δ for juror i if there exists a sequence $W^1, ..., W^s \in 2^N_w$ such that

(i) $W^1 = W$ and $W^s = \hat{W}$.

(ii) for each $q \in \{2, ..., s\}$ there exist $a^q, b^q \in N$ such that

  (a) (W^{q-1}, W^q) are (a^q, b^q) variant.

[3] In the Amorós (2010) model, except for the trivial case in which the planner knows a juror who treats all contestants fairly (i.e. there is $i \in J$ such that $F_i = N$ for each $F \in \Omega$), Condition A cannot be fulfilled if there are only two jurors.

(b) For all $F \in \Omega$, $(a^q, b^q) \in F_i$.

(iii) for each $r \neq q$, $a^q \neq a^r$ and $a^q \neq b^r$.

Property δ is illustrated in Example 4 in the Appendix. We note three important facts about Property δ.
(i) If $(W, \hat{W})$ satisfies Property δ for i, then so does $(\hat{W}, W)$.
(ii) Consider $W, W', W'' \in 2^N_w$. Suppose (W, W') and (W', W'') satisfy Property δ for i. Then (W, W'') satisfies Property δ for i.
(iii) If $(W, \hat{W})$ satisfy Property δ for i, then $W P_i(W) \hat{W}$ and $\hat{W} P_i(\hat{W}) W$.

Parts (i) and (ii) can be easily verified. Part (iii) holds for the following reason. The only difference between any two consecutive sets in the sequence (W^{q-1}, W^q) is that a^q is replaced by b^q. Morover since $W^1 = W$ and $a^q \neq b^r$ for each $r \neq q$, we have $a^q \in W$ and $b^q \notin W$. We know that $(a^q, b^q) \in F_i$ for all $F \in \Omega$ and for all q.[4] Thus $W^1 P_i(W) W^2$, $W^2 P_i(W) W^3$, $\ldots$, $W^{s-1} P_i(W) W^s$. Therefore $W P_i(W) \hat{W}$. Similarly $\hat{W} P_i(\hat{W}) W$.

The next step is to construct a function $\alpha : 2^N_w \times 2^N_w \rightarrow 2^N_w$ satisfying the following properties.

(i) $\alpha(W, W) = W$ for all $W \in 2^N_w$.

(ii) $\alpha(W, \hat{W}) = W$ if $(W, \hat{W})$ satisfies Property δ for i.

(iii) $\alpha(W, \hat{W}) = \hat{W}$ if $(W, \hat{W})$ satisfies Property δ only for j.

(iv) If $(W, \hat{W})$ does not satisfy Property δ for both jurors, then $\alpha(W, \hat{W}) = W_h$ such that

  (a) $W_h \in 2^N_w \setminus \{W, \hat{W}\}$.

  (b) $(\hat{W}, W_h)$ satisfies Property δ for i.

  (c) (W_h, W) satisfies Property δ for j.

Note that if $(W, \hat{W})$ satisfies Property δ for both jurors, then $\alpha(W, \hat{W}) = W$.[5]

LEMMA 4 *Suppose Condition A is satisfied. Then there exists $\alpha : 2^N_w \times 2^N_w \rightarrow 2^N_W$ as defined above.*

Proof: Condition A guarantees that for any given pair, there exists a juror who is known to the planner to be fair over the pair. Let $W, \hat{W}$ be such that $W \neq \hat{W}$. Consider an arbitrary

[4]This means that the planner knows that i is the juror who treats (a^q, b^q) fairly.
[5]This means that all ties are broken in favour of juror i.

sequence of pairs $\{(a^q, b^q)\}$ such that $a^q \in W, a^q \notin \hat{W}$ and $b^q \notin W, b^q \in \hat{W}$.[6] We know that for each q, $(a^q, b^q) \in F_k$ (for all $F \in \Omega$) for some $k \in \{i, j\}$. There are two possibilities. The first possibility is that for each q, we have $(a^q, b^q) \in F_k$ (for all $F \in \Omega$) for some $k \in \{i, j\}$. This implies that either part (ii) or (iii) is true, i.e. $(W, \hat{W})$ satisfies Property δ for i or j or both.

The other possibility is that all elements of the sequence do not belong to the fair set of one juror. We know for every q, there exists some $k \in \{i, j\}$ such that $(a^q, b^q) \in F_k$ (for all $F \in \Omega$). We can partition the elements of the sequence into two sets. The set G_i consists of all pairs that juror i is known to treat fairly (all the q's such that $(a^q, b^q) \in F_i$ for all $F \in \Omega$). The set G_j consists of all pairs for which juror j is known to treat fairly.[7] Thus we can redefine the order in the original sequence as $\{\{(a^q, b^q)\}_{q \in G_i}, \{(a^q, b^q)\}_{q \in G_j}\}$. We denote the new sequence of pairs $\{(a^p, b^p)\}$. There exists an integer K ($1 < K < s$) such that (i) $(a^p, b^p) \in F_i$ (for all $F \in \Omega$) for all $p \leq K$ and (ii) $(a^p, b^p) \in F_j$ (for all $F \in \Omega$) for all $p > K$. We obtain a corresponding sequence of sets $\{(W^{p-1}, W^p)\}$ such that (W^{p-1}, W^p) are (a^p, b^p) variant and $W^1 = \hat{W}$, $W^s = W$. By construction, $(\hat{W}, W^K)$ satisfies Property δ for i and (W^K, W) satisfies Property δ for j. The arguments above show that $\alpha(\cdot)$ is well defined and satisfies conditions (i), (ii), (iii) and (iv). ■

The construction of $\alpha(\cdot)$ is illustrated in Example 5 in the Appendix. We now complete the proof of the Theorem.

Let $\Gamma \equiv (M, g)$ be a mechanism where for each $i \in J$, $M_i = 2^N_w \times 2^N_w \times 2^N_w \times \{Y, N\} \times N$. Note that N is the set of non-negative integers. The outcome function is specified by the following rules:

Rule 1: (1.1) If for each $k \in \{i, j\}$, $m_k = (W, \hat{W}_k, \tilde{W}_k, Y, z_k)$, then $g(m) = W$.
(1.2) If for each $k \in \{i, j\}$, $m_k = (W_k, \hat{W}_k, \tilde{W}_k, Y, z_k)$ where $W_i \neq W_j$, then $g(m) = \alpha(W_i, W_j)$.
Rule 2: If $m_i = (W_i, \hat{W}_i, \tilde{W}_i, Y, z_i)$ and $m_j = (W_j, \hat{W}_j, \tilde{W}_j, N, z_j)$ then

(i) $g(m) = \hat{W}_j$ if $(\hat{W}_j, \alpha(W_i, W_j))$ satisfies Property δ for j.

(ii) Otherwise, $g(m) = \alpha(W_i, W_j)$.

Rule 3: If $m_i = (W_i, \hat{W}_i, \tilde{W}_i, N, z_i)$, $m_j = (W_j, \hat{W}_j, \tilde{W}_j, N, z_j)$ and $z_i > z_j$, then $g(m_i, m_j) = \tilde{W}_i$. Ties are broken in favour of agent i.

We will show that the mechanism Γ Nash implements φ.

[6] The sequence of pairs $\{(a^q, b^q)\}$ (for $q \in \{1, \ldots, s\}$ for some integer s) also satisfies the property that for each q and $r \neq q$, we have $a^q \neq a^r$ and $a^q \neq b^r$. For instance, let $N = \{a, b, c, d, e, f\}$ and $w = 3$. Consider the sets abc and def. Some examples of such a sequence are $\{(a, d), (b, e), (c, f)\}$, $\{(a, e), (b, d), (c, f)\}$, $\{(b, d), (a, f), (c, e)\}$, etc.

[7] Any pair which is treated fairly by both jurors can be arbitrarily assigned to either G_i or G_j.

Claim 1: For each $(R, W^*) \in S(\Omega)$, there exists $m \in N(\Gamma, R, W^*)$ such that $g(m) = W^*$.

Let $(R, W^*) \in S(\Omega)$. Consider message $m = ((W^*, \hat{W}_k, \tilde{W}_k, Y, z_k))_{k\in\{i,j\}}$. Then Rule (1.1) of the mechanism is applicable and $g(m) = W^*$. We argue that $m \in N(\Gamma, R, W^*)$.

Fix $i \in J$. Consider a deviation by agent i to $m_i^{'} = (W_i, W_i^{'}, W_i^{''}, Y, z_i^{'})$ with $W_i \neq W^*$. Then Rule (1.2) of the mechanism is applicable and $g(m_i^{'}, m_j) = \alpha(W_i, W^*)$. Let $\alpha(W_i, W^*) \neq W^*$. There are two possibilities. We know that $\alpha(W_i, W^*) = W_i$ if (W_i, W^*) satisfies Property δ for juror i. Then $W^* P_i(W^*) W_i$ and juror i does not improve his welfare by deviating to $m_i^{'}$. The second case is that $\alpha(F_i, F_j, W_i, W^*) = W_h$, where (W^*, W_h) and (W_h, W_i) are admissible for i and j respectively. This implies that $W^*\ P_i(W^*)\ W_h$ and juror i does not improve his welfare by the deviation.

Consider a deviation by agent i to $m_i^{'} = (W_i, W_i^{'}, W_i^{''}, N, z_i^{'})$. Then Rule 2 of the mechanism is applicable. There are two possibilities.

Case 1: $g(m_i^{'}, m_j) = W_i^{'}$. This case arises when $(W_i^{'}, \alpha(W_i, W^*))$ is admissible for juror i.

(i) If (W_i, W^*) satisfies Property δ for juror i, we have $\alpha(W_i, W^*) = W_i$. Thus $(W_i^{'}, W^*)$ satisfies Property δ for i (as $(W_i^{'}, \alpha(W_i, W^*))$ satisfies Property δ for i).

(ii) $\alpha(W_i, W^*) = W^*$. This implies that $(W_i^{'}, W^*)$ satisfies Property δ for i.

(iii) $\alpha(W_i, W^*) = W_h$ where (W^*, W_h) satisfies Property δ for i and (W_i, W_h) satisfies Property δ for j. Thus $(W_i^{'}, W^*)$ is admissible for juror i.

We have shown that $(W_i^{'}, W^*)$ satisfies Property δ for juror i. Therefore $W^* P_i(W^*) W_i^{'}$ and deviation by i to $m_i^{'}$ is not profitable.

Case 2: $g(m_i^{'}, m_j) = \alpha(W_i, W^*)$.[8] The deviation by i to $m_i^{'}$ is not profitable as $W^* P_i(W^*) \alpha(W_i, W^*)$.

Claim 2: For each $(R, W^*) \in S(\Omega)$ and each $m \in M$ such that $g(m) \neq W^*$,
$m \notin N(\Gamma, R, W^*)$.

Let $(R, W^*) \in S(\Omega)$ and $m \in M$ be such that $g(m) \neq W^*$. There are several cases to consider.

Case 1. Rule 1 applies to m. There are two possibilities.

(1.1) Rule (1.1) applies to m. Then $m_k = (W, \hat{W}_k, \tilde{W}_k, Y, z_k)$ for all $k \in \{i, j\}$.

Since $W \neq W^*$, there exists $(a, b) \in N^2$ such that (i) $a \in W^*$, $a \notin W$ and (ii) $b \in W$, $b \notin W^*$. Let $\hat{W} \in 2^N_w$ be such that $(\hat{W}, W)$ are (a, b) variant.

From Condition A, there exists $k \in \{i, j\}$ such that $(a, b) \in F_k$ for all $F \in \Omega$. Thus $\hat{W} P_k(W^*) W$ and $(W, \hat{W})$ satisfies Property δ for k.

[8]This case arises when $(W_i^{'}, \alpha(W_i, W^*))$ does not satisfy Property δ for juror i.

Suppose $k = i$. Consider a unilateral deviation by agent i to $\hat{m}_i = (\hat{W}, \hat{W}_i, \tilde{W}_i, Y, z_i)$. Rule (1.2) is applicable to the message $(\hat{m}_i, m_j)$ and $g(\hat{m}_i, m_j) = \alpha(\hat{W}, W) = \hat{W}$ (since $(\hat{W}, W)$ satisfies Property δ for i). Thus $m \notin N(\Gamma, R, W^*)$.

Suppose $k = j$ and $k \neq i$. Consider a unilateral deviation by agent j to $\hat{m}_j = (\hat{W}, \hat{W}_j, \tilde{W}_j, Y, z_j)$. Rule (1.2) is applicable and since $k \neq i$, we have $g(m_i, \hat{m}_j) = \alpha(W, \hat{W}) = \hat{W}$. Thus $m \notin N(\Gamma, R, W^*)$.

(1.2) Rule (1.2) applies to m. Then $m_i = (W_i, \hat{W}_i, \tilde{W}_i, Y, z_i)$ and $m_j = (W_j, \hat{W}_j, \tilde{W}_j, Y, z_j)$ with $W_i \neq W_j$. The outcome is $g(m) = \alpha(W_i, W_j) \neq W^*$.

We have to consider the following cases.

(i) $\alpha(W_i, W_j) = W_i$. This is the case when (W_i, W_j) satisfies Property δ for i. Since $g(m) = W_i \neq W^*$, there exists $a, b \in N$ such that (i) $a \in W^*$, $a \notin W_i$ and (ii) $b \in W_i$, $b \notin W^*$. Let $\hat{W} \in 2^N_w$ be such that $(\hat{W}, W_i)$ are (a, b) variant. From Condition A, there exists $k \in \{i, j\}$ such that $(a, b) \in F_k$ for all $F \in \Omega$. We have $\hat{W} P_k(W^*) W_i$ and $(\hat{W}, W_i)$ satisfies Property δ for agent k.

Suppose $k = i$. We know that both (W_i, W_j) and $(W_i, \hat{W})$ satisfy Property δ for i. Thus $(\hat{W}, W_j)$ also satisfies Property δ for agent i. Consider a unilateral deviation by i to $\hat{m}_i = (\hat{W}, \hat{W}_i, \tilde{W}_i, Y, z_i)$. Rule 1 is applicable and $g(\hat{m}_i, m_j) = \hat{W}$. Since $\hat{W} P_i(W^*) W_i$, we have $m \notin N(\Gamma, R, W^*)$.

Suppose $k = j$ and $k \neq i$. Consider a unilateral deviation by j to $\hat{m}_j = (W_j, \hat{W}, \tilde{W}_j, N, z_j)$. Rule 2 is applicable to $(m_i, \hat{m}_j)$ and $g(\hat{m}_i, m_j) = \hat{W}$ (since $(W_i, \hat{W})$ satisfies Property δ for j). Thus $m \notin N(\Gamma, R, W^*)$.

(ii) $\alpha(F_i, F_j, W_i, W_j) = W_j$. This case arises when (W_i, W_j) satisfies Property δ only for j. It can be shown that $m \notin N(\Gamma, R, W^*)$ (using arguments similar to those used in (i)).

(iii) $\alpha(F_i, F_j, W_i, W_j) = W_h$, where (W_j, W_h) satisfies Property δ for i and (W_h, W_i) satisfies Property δ for j. Since $g(m) = W_h \neq W^*$, there exists $(a, b) \in N^2$ such that (i) $a \in W^*$, $a \notin W_h$ and (ii) $b \in W_h$, $b \notin W^*$. Let $\hat{W} \in 2^N_w$ be such that $(\hat{W}, W_h)$ are (a, b) variant. From Condition A, there exists $k \in \{i, j\}$ such that $(a, b) \in F_k$ for all $F \in \Omega$. So $\hat{W} P_k(W^*) W_h$ and $(\hat{W}, W_h)$ satisfies Property δ for agent k. Fix $k = j$. Consider a unilateral deviation by agent j to $\hat{m}_j = ((W_j, \hat{W}, \tilde{W}_j, N, z_j)$. Rule 2 is applicable to $(m_i, \hat{m}_j)$ and $g(m_i, \hat{m}_j) = \hat{W}$. This is because $(\hat{W}, \alpha(W_i, W_j))$ ($\alpha(W_i, W_j) = W_h$) satisfies Property δ for j. Thus $m \notin N(\Gamma, R, W^*)$.

Case 2. Rule 2 applies to m. Then $m_i = (W_i, \hat{W}_i, \tilde{W}_i, Y, z_i)$ and $m_j = (W_j, \hat{W}_j, \tilde{W}_j, N, z_j)$. We refer to agent j as the dissident (the agent who has announced N in the message m). There are several cases to consider.

(2.1) $g(m) = \hat{W}_j$. This case arises when $(\hat{W}_j, \alpha(W_i, W_j))$ satisfies Property δ for j. Since $\hat{W}_j \neq W^*$, there exists $(a, b) \in N^2$ such that (i) $a \in W^*$, $a \notin \hat{W}_j$ and (ii) $b \in \hat{W}_j$, $b \notin W^*$. Let

$\hat{W} \in 2^N_w$ be such that $(\hat{W}, W_j)$ are (a,b) variant. From Condition A, there exists $k \in \{i,j\}$ such that $(a,b) \in F_k$ for all $F \in \Omega$. We have $\hat{W} P_k(W^*) W_j$ and $(\hat{W}, W_j)$ satisfies Property δ for agent k.

Suppose $k = i$. Consider a unilateral deviation by i to $\hat{m}_i = (W_i, \hat{W}_i, \hat{W}, N, \hat{z}_i)$ such that $\hat{z}_i > z_j$. Rule 3 applies to $(\hat{m}_i, m_j)$ and $g(\hat{m}_i, m_j) = \hat{W}$. Thus $m \notin N(\Gamma, R, W^*)$.

Suppose $k = j$. We know that $(\hat{W}, \hat{W}_j)$ and $(\hat{W}_j, \alpha(W_i, W_j))$ both satisfy Property δ for j.Thus $(\hat{W}, \alpha(W_i, W_j))$ satisfies Property δ for j. Consider a unilateral deviation by j to $\hat{m}_j = (W_j, \hat{W}, \tilde{W}_j, N, z_j)$. Rule 2 is applicable and $g(m_i, \hat{m}_j) = \hat{W}$. Since $\hat{W}\ P_j(W^*)\ \hat{W}_j$, we have $m \notin N(\Gamma, R, W^*)$.

(2.2) $g(m) = \alpha(W_i, W_j)$.[9]

(i) $\alpha(W_i, W_j) = W_i$ if (W_i, W_j) satisfies Property δ for i. Since $g(m) = W_i \neq W^*$, there exist $(a,b) \in N^2$ such that (i) $a \in W^*$, $a \notin W_i$ and (ii) $b \in W_i$, $b \notin W^*$. Consider $\hat{W} \in 2^N_w$ such that $(\hat{W}, W_i)$ are (a,b) variant.

From Condition A, there exists $k \in \{i,j\}$ such that $(a,b) \in F_k$ for all $F \in \Omega$. Thus we have $\hat{W} P_k(W^*) W_i$ and $(\hat{W}, W_i)$ satisfies Property δ for agent k.

Suppose $k = j$. Consider a unilateral deviation by j to $\hat{m}_j = (W_j, \hat{W}, \tilde{W}_j, N, z_j)$. Rule 2 applies to $(m_i, \hat{m}_j)$ and $g(m_i, \hat{m}_j) = \hat{W}$ (as $\alpha(W_i, W_j) = W_i$ and $(\hat{W}, \alpha(W_i, W_j))$ satisfies Property δ for j). Since $\hat{W} P_j(W^*) W_i$, we have $m \notin N(\Gamma, R, W^*)$.

Suppose $k = i$. Consider a unilateral deviation by i to $\hat{m}_i = (W_i, \hat{W}_i, \hat{W}, N, \hat{z}_i)$ with $\hat{z}_i > z_j$. Rule 3 is applicable and $g(\hat{m}_i, m_j) = \hat{W}$. Thus $m \notin N(\Gamma, R, W^*)$.

(ii) $\alpha(F_i, F_j, W_i, W_j) = W_j$. It can be shown that $m \in N(\Gamma, R, W^*)$ (using arguments similar to those used in (i)).

(iii) $\alpha(F_i, F_j, W_i, W_j) = W_h$, where $W_h \neq W_i$ and $W_h \neq W_j$. We have $m \notin N(\Gamma, R, W^*)$ (using arguments similar to those used in (i)).

Case 3. Rule 3 applies to m. Then $m_i = (W_i, \hat{W}_i, \tilde{W}_i, N, z_i)$ and $m_j = (W_j, \hat{W}_j, \tilde{W}_j, N, z_j)$. We assume that $z_i > z_j$ and $g(m) = \tilde{W}_i$.

Since $\tilde{W}_i \neq W^*$, there exists $(a,b) \in N^2$ such that (i) $a \in W^*$, $a \notin \tilde{W}_i$ and (ii) $b \in \tilde{W}_i$, $b \notin W^*$. Let $\hat{W} \in 2^N_w$ such that $(\hat{W}, \tilde{W}_i)$ are (a,b) variant. By Condition A, there exists $k \in \{i,j\}$ such that $(a,b) \in F_k$ for all $F \in \Omega$.

Suppose $k = i$. Consider a unilateral deviation by agent i to $\hat{m}_i = (W_i, \hat{W}_i, \hat{W}, N, z_i)$. Rule 3 is applicable and $g(\hat{m}_i, m_j) = \hat{W}$. Since $\hat{W} P_i(W^*) W_i$, we have $m \notin N(\Gamma, R, W^*)$.

Suppose $k = j$. Consider a unilateral deviation by agent j to $\hat{m}_j = (W_j, \hat{W}_j, \hat{W}, N, \hat{z}_j)$ with $\hat{z}_j > z_i$. Rule 3 applies and $g(m_i, \hat{m}_j) = \hat{W}$. Since $\hat{W} P_j(W^*) W_i$, we have $m \notin N(\Gamma, R, W^*)$.

[9] This case arises when $(\hat{W}_j, \alpha(W_i, W_j))$ does not satisfy Property δ for agent j and the dissident's announcement $\hat{W}_j$ is not chosen as the outcome.

Cases 1, 2 and 3 are exhaustive. Therefore Claim 2 is established and the proof is complete. ■

We would like to point out an important difference between our model and that of Amorós (2010) that has an important bearing on the two person case. In our model, fairness is defined over individual pairs of contestants, i.e $F_i \subseteq N^2$. It is therefore possible for a juror to be fair over (a, b) and (b, c) without being fair over (a, c). In Amorós (2010) model, jurors are fair over *subsets* of contestants, i.e. $F_i \subseteq N$. In this case, a juror who is fair over (a, b) and (b, c) has to be fair over (a, c). It follows that if Condition A is satisfied, there must exist a juror who is known to be fair over *all pairs*. For such a juror, the maximal element in any preference ordering is the true set of deserving winnners. The implementation problem is now trivial - the dictatorial mechanism where this juror (Condition A requires the identity of this juror to be known to the planner) is the dictator will always implement φ. On the other hand, in our model, Condition A can be satisfied without the existence of a juror who is known to the planner to be fair over all pairs. For instance, $N = \{a, b, c\}$, $(a, b), (b, c) \in F_1$ and $(a, c) \in F_2$ for all $F \in \Omega$ satisfies Condition A. The implementation problem is no longer trivial. In Theorem 4, we show that Condition A is necessary and sufficient for implementation.

We now show that assuming partial honesty does not enhance implementation possibilities.

Corollary 1 *Assume each juror i is F_i-partially honest. Condition A is necessary and sufficient for implementation.*

Proof: The arguments in Theorem 3 apply to this case without any change. Suppose Condition A does not hold. It follows that for each pair there exist two jurors who treat the pair fairly in each $F \in \Omega$. Thus both jurors are known to treat all pairs fairly and Condition A holds. Therefore Condition A always holds.

The mechanism used in the sufficiency part of Theorem 4 can be used once again in this case. It suffices to note that truth-telling will continue to remain an equilibrium when players are partially honest. ■

The results in two person case stand in contrast to the many person implementation.

3.6 Fairness and partial honesty

We have assumed that a juror can be partially honest over a contestant pair only if he is fair over that pair. In this section, we show that partial honesty cannot substitute for fairness. Suppose we redefine partial honesty to apply to *all pairs* irrespective of whether or not the juror is fair over the pair. Definition 3 can be suitably qualified to include all pairs of contestants (replace for all $(a, b) \in F_i$ by $(a, b) \in N^2$).

Implementation will not be possible if there exists a pair for which no juror is fair. We illustrate this with an example.

EXAMPLE **3** Let $N = \{a, b, c\}$, $|J| = 3$ and $w = 1$. Let $\Omega = \{F, \ldots\}$. Let $F_1 = \{(b, c)\}$ and $F_2 = F_3 = \{(a, c)\}$. Note that $(a, b) \notin F_i$ for all $i \in J$. Consider R below, which is admissible with respect to F.

	R_1			R_2			R_3		
$W^* =$	a	b	c	a	b	c	a	b	c
	a	a	a	a	a	b	a	a	b
Pref.	b	b	c	b	b	c	b	b	c
	c	c	b	c	c	a	c	c	a

Table 3.6: Example 3

Note that $R_i(a) = R_i(b)$ for all $i \in J$.

Suppose that φ is implementable by the mechanism $\Gamma = (M, g)$. By definition, there exists $m \in M$ such that $m \in N(\Gamma, R, a)$ and $g(m) = a$. Since $g(m) \neq b$, it must be the case that $m \notin N(\Gamma, R, b)$. Since a is $R_i(b)$-maximal for all i, there must exist a juror i (say 1) and a message $m_1^{'}$ such that $g(m_1^{'}, m_{-1}) = a$ and $m_1^{'} = \{b\} \times \cdots$, i.e. $m_1^{'}$ involves the announcement of the true winner b by 1. Now consider the message $(m_1^{'}, m_{-1})$. Since $g(m_1^{'}, m_{-1}) = a$ and a is $R_i(b)$-maximal for all i, there must exist a juror $j \in \{2, 3\}$ (say 2) and a message $m_2^{'}$ such that $g(m_1^{'}, m_2^{'}, m_3) = a$ and $m_2^{'} = \{b\} \times \cdots$, i.e. $m_2^{'}$ involves the announcement of the true winner b by 2. Repeating the argument, there must exist a message $m_3^{'}$ such that $g(m_1^{'}, m_2^{'}, m_3^{'}) = a$ and $m_3^{'} = \{b\} \times \cdots$. However note that $(m_1^{'}, m_2^{'}, m_3^{'}) \in N(\Gamma, R, b)$ since the outcome is a and all agents are announcing the true winner b. This contradicts the assumption that φ is implementable.

It is clear that the necessity of jurors who are fair over any given contestant pair holds generally. We therefore see partial honesty as a strengthening of the fairness condition rather than a replacement for it.

3.7 Conclusion

The paper explores the consequences of introducing the behaviourial assumption of partial honesty for jurors in the model proposed by Amorós (2010). We show that this assumption leads to a weakening of the conditions for implementation in the case of three or more agents. However no such weakening is possible in the case of two agents.

3.8 Appendix

Examples 4 and 5 below illustrate various features of the construction if the mechanism used in proof of Theorem 4.

Example **4** Let $N = \{a, b, c, d\}$, $w = 2$ and $2^N_w = \{ab, ac, ad, bc, bd, cd\}$. The sets ab and cd differ in the elements $\{a, b, c, d\}$. Suppose the structure Ω satisfies either (1) or (2).

1. $(a, c), (b, d) \in F_i$ for all $F \in \Omega$.

2. $(a, d), (b, c) \in F_i$ for all $F \in \Omega$.

Assume that (1) holds. Consider the sequence $\{ab, cb, cd\}$. This sequence of sets is such that (i) (ab, cb) are (a, c) variant where $(a, c) \in F_i$ (for all $F \in \Omega$) and (ii) (cb, cd) are (b, d) variant where $(b, d) \in F_i$ (for all $F \in \Omega$). Thus (ab, cd) satisfy Property δ for juror i. We have $abP_i(ab)cb$ (as $(a, c) \in F_i$) and $cbP_i(ab)cd$ (as $(b, d) \in F_i$). Thus $abP_i(ab)cd$.

Example **5** Let $N = \{a, b, c, d\}$, $w = 2$ and $2^N_w = \{ab, ac, ad, bc, bd, cd\}$. Suppose Ω is such that $F_i = \{(a, b), (a, c), (a, d)\}$ and $F_j = \{(b, c), (b, d), (c, d)\}$ for all $F \in \Omega$. Consider the sets $W = ab$ and $\hat{W} = cd$. The possible sequences between ab and cd are $\{(a, c), (b, d)\}$, $\{(a, d), (b, c)\}$, $\{(b, d), (a, c)\}$ and $\{(a, d), (b, c)\}$. Note that (ab, cd) is not admissible for both jurors. Thus $\alpha(ab, cd) \notin \{ab, cd\}$.

Consider the sequence $\{(a, c), (b, d)\}$. Note that (cd, ad) is admissible for juror i as $(a, c) \in F_i$ (for all $F \in \Omega$). Also (ad, ab) is admissible for juror j. For the sequence $\{(a, d), (b, c)\}$, we know that $(a, d) \in F_i$ and $(b, c) \in F_j$ (for all $F \in \Omega$). We have (cd, ac) and (ac, ab) are admissible for i and j respectively. Thus $\alpha(ab, cd) \in \{ad, ac\}$.

Chapter 4

The equivalence between adjacent non-manipulability and strategy-proofness in voting domains: A sufficiency result

4.1 Introduction

In any model where the agents have private information, the primary objective of the mechanism designer is to design rules which provide agents the incentive to reveal their private information truthfully. Incentive compatibility guarantees that every agent truthfully reveals his private information, irrespective of the announcements made by other agents. Incentive compatibility assumes that every feasible preference is a candidate for manipulation. Thus the task of designing truthful rules may be too demanding in many settings, as it requires the mechanism designer to check all possible incentive constraints.

On the other hand if the rule is immune to candidate manipulations that are "near" or "close" to the true preference of an agent, then the rule satisfies local incentive compatibility. Every incentive compatible rule is locally incentive compatibilty. However it is possible that a rule is locally incentive compatible, but not incentive compatible. The question that we are interested in is when a locally incentive compatible rule is also incentive comptaible. In any setting where local incentive compatibility implies incentive compatibility, the designer's task is now only involves checking only the local incentive constraints.

Several papers have examined this issue in different settings. Sato (2013b) considers the standard voting model without transfers. He provides a sufficient condition and a weaker necessary condition for the equivalence. Sato (2013a) shows that the results obtained about the equivalence do not carry forward to the case where the domain includes weak orders.

We prove a sufficiency result that is weaker than Sato's sufficiency result.

In related literature, Carroll (2012) considers both voting models and allocation models with transfers. In voting models, he shows that a number of specific domains such as the full domain, the domain of all single peaked preferences etc are domains where local incentive compatibility implies incentive compatibility. However, he does not provide a general condition for voting domains. In models with transfers, he shows that convexity of the type space is sufficient and almost necessary. Mishra et al. (2015) consider allocation models with transfers. They show that convexity of the type space is not necessary, if an additional assumption on transfers is made. Archer and Kleinberg (2014) examine an allocation model with money and a different notion of local incentive compatibility.

The paper is organized as follows. Section 4.2 gives basic notation and definitions. Section 4.3 describes the existing results. Section 4.4 gives our result and Section 4.5 concludes.

4.2 Basic Notation and Definations

Through out the paper, we shall assume that there is a single agent/voter. In our analysis there is no loss of generality in making this assumption.[1] We let A be a finite set of alternatives with $|A| = m$. The set of linear or antisymmetric orders over the elements of A is denoted by L. Elements of L will be denoted by P, P' etc and will be referred to as preference orderings or orderings. For all $a, b \in A$, aPb is interpreted as "a is strictly preferred to b according to P". For every $P \in L$ and $k \in \{1, \ldots, m\}$, $r_k(P)$ denote the kth ranked alternative in P i.e. $r_k(P) = a \implies |\{b \in A : bPa\}| = k - 1$.

Two alternatives x and y are *contiguous* in P if there exists $k \in \{1, \ldots, m-1\}$ such that $r_k(P) = x$, $r_{k+1}(P) = y$ or $r_k(P) = y$, $r_{k+1}(P) = x$. Two preferences P and P' are adjacent if P' can be obtained by swapping contiguous alternatives x and y in P without changing the ranks of other alternatives different from x and y. If P and P' are adjacent, we shall let $A(P, P')$ be the ordered pair of alternatives which are swapped in P to obtain P' i.e. $A(P, P') = (x, y)$ implies $r_k(P) = x$, $r_{k+1}(P) = y$ and $r_k(P') = y$, $r_{k+1}(P') = x$. Let $A(P)$ denote the set of all preference relations that are adjacent to P.

A *domain* is a set D where $D \subset L$. A domain D is interpreted as the set of admissible preference orderings. A rule (or social choice function) is a map $f : D \to A$.

The standard notion of incentive compatibility is strategy-proofness.

Definition **13** *A rule $f : D \to A$ satisfies strategy-proofness if for every $P, P' \in D$, either $f(P) = f(P')$ or $f(P)Pf(P')$.*

If a rule is strategy-proof, an agent cannot manipulate, i.e. cannot get a strictly preferred alternative by misrepresenting her true preference.

[1] Equivalently we can assume that the preferences of all other agents is fixed. For example see Mishra et al. (2015).

A local notion of incentive compatibility introduced by Sato (2013b) and Carroll (2012) is *AM-proofness* (Adjacency Manipulable proofness).

Definition 14 *A rule $f : D \to A$ satisfies AM-proofness if for every $P \in D$ and every $P' \in A(P) \cap D$, either $f(P) = f(P')$ or $f(P)Pf(P')$.*

The AM-proofness of a rule can be characterized by some elementary properties. We describe these below using the terminology of Gibbard (1977).

Definition 15 *The rule $f : D \to A$ is local and non-perverse if for every $P, P' \in D$ with $P' \in A(P) \cap D$ and $A(P, P') = (x, y)$ we have*

(i) $[f(P) = y] \implies [f(P') = y]$

(ii) $[f(P) = x] \implies [f(P') \in \{x, y\}]$

(iii) $[f(P) = z] \implies [f(P') = z]$ *when* $z \neq x, y$

This definition is illustrated in Table 4.1.

P	P'	P	P'	P	P'
z	z	z	z	$[z]$	$[z]$
.	.	.	.	.	.
x	$[y]$	$[x]$	$[y]$	x	y
$[y]$	x	y	$[x]$	y	x
.	.	.	.	.	.

Table 4.1: AM-proofness

Proposition 1 *A rule f is AM-proof iff it is local and non-perverse.*

The proof of this standard and may be found in Sato (2013b) .

It is clear that a strategy-proof rule is also AM-proof. The goal of this paper is to analyze domains where the converse holds.

Definition 16 *A domain D satisfies equivalence if every AM-proof rule is also strategy-proof.*

The next section discusses existing results in this area.

4.3 Existing Results

Carroll (2012) shows that several well known domains satisfy the equivalence for both deterministic and random rules.

Sato (2013b) investigates the question more generally. He gives a sufficient condition on domains that ensures the equivalence. He also provides a weaker necessary condition. We illustrate the issues involved and the results in Sato (2013b). The example below illustrates that not every domain satisfies equivalence.

Example **6** Let $A = \{a, b, c\}$ and D be the following domain.

P^1	P^2	P^3
a	$[b]$	c
$[b]$	a	b
c	c	$[a]$

Table 4.2: Domain D

P^3 is not adjacent to any preference in D. Consider the following rule;

$$f(P) = \begin{cases} b & \text{if } P \in \{P^1, P^2\} \\ a & \text{otherwise} \end{cases}$$

Note that f is AM-proof since no restrictions are imposed on $f(P^3)$. However f is not strategy-proof since $f(P^3)P^1f(P^1)$.

Definition **17** *A path from P to P' in D, denoted by $\sigma(P, P')$, is a sequence of distinct preferences $(P^1, P^2, \ldots, P^L)$ in D satisfying*

(i) $P = P^1$ *and* $P' = P^L$.

(ii) $P^{h+1} \in A(P^h)$, $h \in \{1, 2, \ldots, L-1\}$.

Let $\sum(P, P')$ denote the set of all paths between P and P' in D. Two preferences $P, P' \in D$ are connected in D if there exists a path between P and P' in D. A domain D is connected if for all $P, P' \in D$, there exists a path between P and P' in D. The domain D in Example 6 is not connected.

Proposition 2 (Sato (2013b)) *If Domain D satisfies equivalence then D is connected.*

P^1	P^2	P^3	P^4	P^5
x	$[y]$	$[y]$	$[y]$	$[x]$
$[y]$	x	x	x	y
v	v	v	z	z
w	w	z	v	v
z	z	w	w	w

Table 4.3: Domain $D^{'}$

Connectedness is necessary for the equivalence of AM-proofness and strategy-proofness on a domain. However connectedness is not sufficient for equivalence. Example 7 below illustrates this i.e. the domain D' is connected but does not satisfy equivalence.

Example **7** Let $A = \{x, y, v, w, z\}$ and D' be the following domain.

Let the rule f be defined as follows;

$$f(P) = \begin{cases} x & \text{if } P = P^1 \\ y & \text{otherwise} \end{cases}$$

Proposition 1 can be applied to verify that f is AM-proof. The outcome at P^1 is y; the contiguous alternatives x and y are swapped from P^1 to P^2. Applying Proposition 1, the outcome at P^2 is y. Since there are no contiguous swaps involving y from P^2 to P^3 and from P^3 to P^4, the outcome remains y (by part (iii)). Since y and x are contiguous and are swapped between P^4 and P^5, according to part (ii) of Proposition 1, the outcome at P^5 is either x or y. However if $f(P^5) = x$ then strategy-proofness is violated because $f(P^5)P^1f(P^1)$.

Example 7 identifies the reason why equivalence fails in a connected domain. There may exist alternatives x and y that are contiguous and swapped initially in the sequence and reverse swapped later in the sequence. The lack of symmetry between conditions (i) and (ii) in Proposition 1 can lead to a failure of strategy-proofness. We shall refer to the path $(P^1, P^2, P^3, P^4, P^5)$ as a *problem path*. Strategy-proofness can be restored if there exists an alternative *antidote path* that rules out $f(P^1) = y$ and $f(P^5) = x$ in Example 7. This is illustrated in Example 8 below.

Example **8** Consider the domain D'' below.

Note that D'' consists of D' with the additional ordering P^6.

Consider the rule f defined in Example 7. We claim that there is no AM-proof extension of f on the domain D''. To see this observe $f(P^5) = x$ and Proposition 1 imply $f(P^6) = f(P^7) = f(P^8) = f(P^1) = x$ contradicting $f(P^1) = y$.

P^1	P^2	P^3	P^4	P^5	P^6
x	y	y	y	x	x
y	x	x	x	y	y
v	v	v	z	z	v
w	w	z	v	v	z
z	z	w	w	w	w

Table 4.4: Domain D''

The path (P^1, P^6, P^5) in Example 7 is the antidote path and it addresses the potential difficulties arising from the problem path $(P^1, P^2, P^3, P^4, P^5)$. The insights from Examples 7 and 8 are formalized in the sufficiency result of Sato (2013b) which we state below.

Fix $a, b \in A$. The path $(P^1, \ldots, P^L)$ is with $\{a, b\}$ *restoration* if $A(P^k, P^{k+1}) = (a, b)$ and $A(P^h, P^{h+1}) = (b, a)$ for some distinct $h, k \in \{1, \ldots, L-1\}$ or $A(P^k, P^{k+1}) = (b, a)$ and $A(P^h, P^{h+1}) = (a, b)$. In Example 7, the path $(P^1, P^2, P^3, P^4, P^5)$ is with $\{x, y\}$ restoration.

Theorem **5** (**Sato** (**2013b**)) *Suppose that for each $P, P' \in D$, there exists a path $(P^1, \ldots, P^L)$ in D from P to P' which satisfies the following property: if there exists $x, y \in A$ such that the path $(P^1, \ldots, P^L)$ is with $\{x, y\}$ restoration and xPy, then for each $h \in \{1, \ldots, L\}$ such that yP^hx and $xP^{h+1}y$, there exists a path from P to P^{h+1} along which x overtakes no alternative.*

We know that for any two adjacent preferences P, P', $A(P, P')$ is the ordered pair of alternatives which are swapped in P to obtain P. Let $A(P, P') = (x, y)$. Then we say that x overtakes y in the passage from P to P'. Any path along which x overtakes no alternatives is such that it does not contain the contiguous swap (x, q) for any $q \in A \setminus \{x\}$.

In Example 8, x does not overtake any alternative in the path (P^1, P^6, P^5). Sato (2013b) also proved the following necessary condition.

Proposition 3 (Sato (2013b)) *If D satisfies equivalence then for each $P, P' \in D$ and each $x, y \in A$, there exists a path in D between P and P' which is without $\{x, y\}$ restoration.*

4.4 The Result

In this section we prove a sufficiency result that is weaker than Sato's sufficiency result. The idea behind the new sufficient condition is quite simple. Consider the path in Example 7 which we referred to as a problem path. An alternative antidote path was required which ruled out the case where $f(P) = y$ and $f(P') = x$. In the antidote path from P' to P, x is always "rising" in preference orderings which rules out $f(P) = y$. However the antidote path can be constructed without this condition being satisfied. Suppose there exists a path between P and P' without a sequence of contiguous swaps that begin with y and end in x; clearly $f(P) = y$ and $f(P') = x$ is ruled out.

We define the condition precisely below. We illustrate all concepts and definitions by means of Example 9.

Fix $\sigma(P, P') = (P^1, \ldots, P^L) \in \sum(P, P')$. The path $\sigma(P, P')$ is associated with a sequence of ordered pairs of alternatives,

$$S(\sigma(P, P')) = \{A(P^s, P^{s+1}) = (u^s, u^{s+1}) : s \in \{1, \ldots, L-1\}\}$$

EXAMPLE **9** Let $A = \{x, y, v, z, w, u\}$ and $\hat{D}$ be the following domain. Fix alternatives $x, y \in A$ and preferences $P^1, P^{10} \in \hat{D}$. There is only one path between P^1 and P^{10}. So $\sum(P^1, P^{10}) = \{(P^1, P^2, P^3, P^4, P^5, P^6, P^7, P^8, P^9, P^{10})\}$.

P^1	P^2	P^3	P^4	P^5	P^6	P^7	P^8	P^9	P^{10}
x	x	v	v	v	v	v	v	x	x
y	v	x	x	z	z	x	x	v	y
v	y	y	z	x	x	z	y	y	v
z	z	z	y	y	y	y	z	z	z
w	w	w	w	w	u	u	u	u	u
u	u	u	u	u	w	w	w	w	w

Table 4.5: Domain $\hat{D}$

In Example 9, $S(P^1, P^2, \ldots, P^{10})$ is described below.

$$S((P^1, P^2, \ldots, P^{10})) = \{(y, v), (x, v), (y, z), (x, z), (w, u), (z, x), (z, y), (v, x), (v, y)\}$$

Observe that (y, v) is the contiguous swap between P^1 and P^2, (x, v) the contiguous swap between P^2 and P^3, (y, z) the contiguous swap between P^3 and P^4 and so on.

Fix $a, b \in A$. An *essential sequence* between the alternatives a, b on a path is a sequence of contiguous swaps that begin with a and end in b on this path. Formally $\sigma^e(\sigma(P, P'); a, b)$ is an ordered selection $\{(u^{s_i}, u^{s_{i+1}}) : i \in \{1, \ldots, H\}\}$ from $S(\sigma(P, P'))$ such that (i) $u^{s_1} = a$ and $u^{s_H} = b$ and (ii) $u^{s_i+1} = u^{s_{i+1}}$ for all s_i. The set of all such essential sequences is denoted by $\sum^e(\sigma(P, P'); a, b)$.

In Example 9, observe that $\{(x, v), (v, y)\}$ is an essential sequence since it is a selection from $S((P^1, P^2, \ldots, P^{10})$ that begins with x and ends with y. Similarly $\{(x, z), (z, y)\}$ is also an essential sequence. Therefore

$$\sum^e((P^1, P^2, \ldots, P^{10}); x, y) = \{\{(x, v), (v, y)\}, \{(x, z), (z, y)\}\}.$$

DEFINITION **18** *(Condition α) Let $P^1, P^h, P^{h+1} \in D$ and $x, y \in A$ be such that xP^1y, $P^{h+1} \in A(P^h)$ and $A(P^h, P^{h+1}) = (y, x)$. Let $Z(x, y) = \{z \in A \setminus \{x, y\} : xP^1z, zP^1y$ and $yP^{h+1}z\}$. For all $w \in Z(x, y) \cup \{y\}$, there exists a path $\sigma \in \sum(P, P^{h+1})$ such that $\sum^e(\sigma; w, x) = \emptyset$ and $\sum^e(\sigma; x, w) = \emptyset$.*

This condition is satisfied if there exists an alternative path without an essential sequence for every problem path. The formal proof for the sufficiency of Condition α is provided below.

THEOREM **6** *If D is connected and satisfies Condition α, it satisfies equivalence.*

The proof uses the following lemma.

LEMMA **5** *Fix $a, b \in A$ and preferences $P, P' \in D$ such that $[aPb$ and $aP'b]$. Suppose there exists $\sigma \in \sum(P, P')$ such that $\sum^e(\sigma; b, a) = \emptyset$. Then $[f(P) = b] \implies [f(P') \neq a]$.*

Proof: Suppose the claim is false, i.e $f(P) = b$ and $f(P') = a$. Let $\sigma(P, P') = (P^1, P^2, \ldots, P^l)$.

Since $f(P') \neq b$ and f is AM-proof, there exists $k_1 \in \{1, \ldots, l-1\}$ such that $A(P^{k_1}, P^{k_1+1}) = (b, u_1)$ for some $u_1 \in A \setminus \{b\}$. Proposition 1 implies that $f(P^{k_1+1}) = u_1$. If $u_1 = a$ then $(b, a) \in \sum^e(\sigma(P, P'); b, a)$, which is in contradiction with the assumption that $\sum^e(\sigma(P, P'); b, a) = \emptyset$. If $u_1 \neq a$ (note that $u_1 \in A \setminus \{b, a\}$) then AM-proofness of f and $f(P') = a$ imply that: there exists $k_2 \in \{k_1 + 1, \ldots, l-1\}$ and $u_2 \in A \setminus \{b, u_1\}$ such that $A(P^{k_2}, P^{k_2+1}) = (u_1, u_2)$. This follows from Proposition 1 and the fact that $f(P^{k_1+1}) \neq f(P')$. Proposition 1 also implies that $f(P^{k_2+1}) = u_2$. If $u_2 = a = f(P')$ then $(b, u_1), (u_1, a) \in \sum^e(\sigma(P, P'); b, a)$ which results in a contradiction. However if $u_2 \neq a$, then there exists $k_3 \in \{k_2+1, \ldots, l-1\}$ and $u_3 \in A \setminus \{b, u_1, u_2\}$ such that $A(P^{k_3}, P^{k_3+1}) = (u_2, u_3)$.[2]

In this manner, at each step we obtain k_i and the corresponding alternative u_i such that $A(P^i, P^{i+1}) = (u_{i-1}, u_i)$. Since $S(\sigma)$ is a finite set, thus there exists $K < l-1$ such that $u_K = a$. This implies that $\{(b, u_1), (u_1, u_2), \ldots, (u_K - 1, a)\} \in \sum^e(\sigma(P, P'); b, a)$, which contradicts our assumption that this set is null. ■

We now begin the proof of the theorem.

Proof: Let D be a connected domain satisfying Condition α and f be an AM-proof rule on D. It is sufficient to show that f is strategy-proof. Let $P, P' \in D$. The lower contour set for alternative x at the preference P is denoted by $L(P, x) = \{y \in A : xPy\} \cup \{x\}$. Our objective is to prove that: $f(P') \in L(P, f(P))$.

[2]Note that u_1, u_2 and u_3 are distinct from each other.

Because D is connected, there exists a path between P and P' in D, $(P^1, \ldots, P^L)$ where $P^1 = P$ and $P^L = P'$.

Step 1: At each step from P^h to P^{h+1} ($h \in \{1, \ldots, L-1\}$) in the path $(P^1, \ldots, P^l)$, the outcome can change only when $A(P^h, P^{h+1}) = (f(P^h), a)$ where a is the alternative right below $f(P^h)$ in P^h. In all other cases the outcome does not change (By Proposition 1).

Step 2: For each $k \in \{1, \ldots, l-1\}$, if $\left[f(P^k) \in L(P, f(P)) \text{ and } f(P) \in L(P^k, f(P^k))\right]$ then $\left[f(P^{k+1}) \in L(P, f(P)) \text{ and } f(P) \in L(P^{k+1}, f(P^{k+1}))\right]$.

By Step 1, it suffices to consider the case where a is the alternative right below $f(P^k)$ in P^k and $A(P^k, P^{k+1}) = (f(P^k), a)$. By AM-proofness, $f(P^{k+1}) \in \{f(P^k), a\}$.

Case 1: $f(P^{k+1}) = a$.

Since $f(P) \in L(P^k, f(P^k))$ and $A(P^k, P^{k+1}) = (f(P^k), x)$. Thus we have $f(P) \in L(P^{k+1}, f(P^{k+1}))$. We claim that $f(P^{k+1}) \in L(P, f(P))$.

Suppose the claim is false i.e. $aPf(P)$. Since $f(P)Pf(P^k)$, this implies that $aPf(P^h)$.

We note that P, P^k, P^{k+1} and the alternatives $x, f(P^h) \in X$ are such that

1. $P^{k+1} \in A(P^k)$.
2. $A(P^k, P^{k+1}) = (f(P^k, a)$.
3. $f(P) \in Z(a, f(P^k))$.

Thus comparing with Condition α, we deduce that $P = P^1$, $P^k = P^h$ and $P^{k+1} = P^{h+1}$. Also $a = x$ and $f(P^k) = y$.

Case 1.1: $f(P) = f(P^k)$. Since the domain satisfies Condition α, there exists $\sigma \in \sum(P, P^{k+1})$ such that $\sum^e(\sigma; f(P^k), a) = \emptyset$. This along with Lemma 5 leads to a contradiction.

Case 1.2: $f(P) \neq f(P^k)$. Since the domain satisfies Condition α, there exists $\sigma' \in \sum(P, P^{k+1})$ such that $\sum^e(\sigma'; f(P), a) = \emptyset$. This along with Lemma 5 leads to a contradiction.

Case 2: $f(P^{k+1}) = f(P^k)$ (where $f(P^k) \neq a$). In this case, it trivially follows that $f(P^{k+1}) \in L(P, f(P))$. Now, $L(P^k, f(P^k)) = L(P^{k+1}, f(P^{k+1})) \cup \{a\}$. To complete the proof, we need to show that $f(P) \neq a$.

Let us suppose by way of contradiction that $f(P) = a$. We have assumed that the outcome at P^{k+1} is $f(P^k)$.

Comparing with Condition α, we know the following:

1. $x = f(P) = a$
2. $y = f(P^k$
3. $P^1 = P$, $P^h = P^k$ and $P^{h+1} = P^{k+1}$.

4. $aPf(P^k)$ and $A(P^k, P^{k+1}) = (f(P^k, a)$.

Thus there exists a path $\sigma \in \Sigma(P, P^{k+1})$ such that $\Sigma^e(\sigma, a, f(P^k)) = \emptyset$. This observation along with Lemma 5 leads to a contradiction.

■

The domain in Example 10 below is connected and satisfies Condition α and thereby admits equivalence.

EXAMPLE **10** Let $X = \{a, x, y, z, v, w, u\}$ and D^s be the following domain.

P^1	P^2	P^3	P^4	P^5	P^6	P^7	P^8	P^9	P^{10}
a	a	a	a	a	x	x	x	x	a
x	y	y	y	x	a	a	a	a	x
y	x	x	x	y	y	y	z	z	z
z	z	z	z	z	z	z	y	y	y
v	v	v	u	u	v	v	v	u	u
w	w	u	v	v	w	u	u	v	v
u	u	w	w	w	u	w	w	w	w

Table 4.6: Domain D^s

It can be verified that the only problem path in this example is $(P^1, P^2, P^3, P^4, P^5)$. Consider preferences P^1, P^4 and P^5. In this example xP^1y, $A(P^4, P^5) = (y, x)$ and $Z(x, y) = \emptyset$. The path $(P^1, P^6, P^7, P^8, P^9, P^{10}, P^5)$ does not contain any essential sequence for the alternatives y, x. This path does not contain any essential sequence for the alternatives x, y. This path is the alternative antidote path which eliminates the non strategy-proof rules generated by the problem path $(P^1, P^2, P^3, P^4, P^5)$. Therefore Condition α is satisfied. Table 4.6 lists the preferences in D^s and Figure 4.1 represents the adjacency relations in D^s.

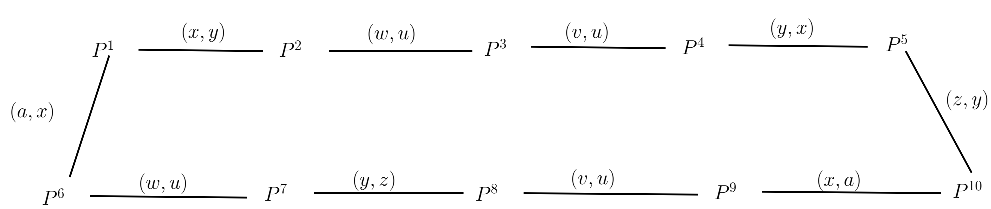

Figure 4.1: Adjacency between preference relations in D^s

We note that a domain satisfying Condition α also satisfies Sato (2013b) sufficient condition.

Proposition 4 *Suppose that for each $P, P' \in D$, there exists a path $(P^1, \ldots, P^L)$ in D from P to P' which satisfies the following property: if there exists $x, y \in A$ such that the path $(P^1, \ldots, P^L)$ is with $\{x, y\}$ restoration and xPy, then for each $h \in \{1, \ldots, L\}$ such that yP^hx and $xP^{h+1}y$, there exists a path from P to P^{h+1} along which x overtakes no alternative. Then D satisfies Condition α.*

Proof: Consider preferences P^1, P^h, and P^{h+1} on the path $(P^1, \ldots, P^L)$. We know xP^1y, $P^{h+1} \in A(P^h)$, and $A(P^h, P^{h+1}) = (y, x)$. Thus there exists an alternative path σ from P^1 to P^{h+1} along which x does not overtake any alternative, i.e. this path does not contain the contiguous swap (q, x) for any $q \in A \setminus \{x\}$.

Claim 1: $\sum^e(\sigma; w, x) = \emptyset$ for all $w \in Z(x, y) \cup \{y\}$ for the path σ.

To see this, consider any $w \in Z(x, y) \cup \{y\}$. Suppose $\sum^e(\sigma; w, x) \neq \emptyset$ and $\{(w, u^1), (u^1, u^2), \ldots, (u^K, x)\}$ belongs to this set. However (u^K, x) contradicts the assumption that σ is a path along which x does not overtake any alternative.

Observation 1: Claim 1 and $[xPw$ and $xP^{h+1}w]$ together imply that for any preference $\hat{P}$ which belongs to the path σ, $x\hat{P}w$.

Suppose not, i.e. there exists $\hat{P}$ on σ where $w\hat{P}x$. This implies that there is a (x, w) swap on σ (between P and $\hat{P}$). We know that xPw, $w\hat{P}x$ and $xP^{h+1}w$. Thus there exists a (w, x) swap on σ (between $\hat{P}$ and P^{h+1}). Thus $(w, x) \in \sum^e(\sigma; w, x)$. This contradicts Claim 1.

Claim 2: $\sum^e(\sigma; x, w) = \emptyset$ for all $w \in Z(x, y) \cup \{y\}$ for the path σ.

To see this, consider any $w \in Z(x, y) \cup \{y\}$. Suppose $\sum^e(\sigma; x, w) \neq \emptyset$ and $\{(x, q_1), (q_1, q_2), \ldots, (q^K, w)\}$ belongs to this set.

(i) Since (x, q_1) belongs to σ, there exists $\hat{P}^1$ on the path σ where $q_1\hat{P}^1x$. We now show that for any preference $\bar{P}$ lying after $\hat{P}^1$ on the path σ, we have $q_1\bar{P}x$. Suppose not i.e. $x\bar{P}q_1$. Since $q_1\hat{P}^1x$, there exists a (q_1, x) swap on σ. This is not possible as by assumption, x does not overtake any alternative on the path σ.

(ii) (q_1, q_2) belongs to the path σ (in particular, it appears after the (x, q_1) swap). Thus there exists a preference $\hat{P}^2$ on the path σ ($\hat{P}^2$ lies after $\hat{P}^1$ on σ), where $q_2\hat{P}^2q_1$ and $q_1\hat{P}^2x$ (from (i)). Note that for any preference $\bar{P}$ after $\hat{P}^2$ on σ is such that $q_2\bar{P}x$.

By following the sequence $\{(x, q_1), (q_1, q_2), \ldots, (q_{K-1}, q_K), (q_K, w)\}$ and using similar arguments, we know that there exists a preference $\hat{P}^K$ on σ where $q_K\hat{P}^Kq_{K-1}$ and $q_{K-1}\hat{P}^Kx$. For any preference $\bar{P}$ lying after $\hat{P}^K$ on σ, we have $q_K\bar{P}x$.

Table 4.7 illustrates the preferences in Claim 2.

The final swap in the sequence present on σ is (q^K, w). We know that any preference $\bar{P}$ after $\hat{P}^K$ will satisfy:

P	$\hat{P}^1$	$\hat{P}^2$	$\hat{P}^K$	P^{h+1}
.	.	.	.	.
x	.	.	.	.
.	q_1	.	q_K	x
.	x	q_2	q_{K-1}	.
w	.	q_1	.	w
.	w	.	x	.
.	.	x	.	.
.	.	.	w	.
.	.	w	.	.

Table 4.7: Preferences in Claim 2

1. $q^K \bar{P} x$
2. $x \bar{P} w$

In order to finally have a (q_K, w) swap on σ and given the placement of the alternatives q_K, x and w in $\hat{P}^K$. There must exist either a (q_K, x) or a (x, w) swap before (q^K, x) on σ.

If there exists (q_1, x), then (i) above is violated. We know that (i) must hold since σ is a path on which x does not overtake any alternative.

If there exists a (x, w) swap on σ, then we have a contradiction by Observation 1.

■

Our result therefore implies Sato (2013b) sufficiency result. Our result is strictly stronger. The domain in Example 10 satisfies Condition α and therefore satisfies equivalence. However we claim that it does not satisfy the sufficiency condition is Sato.

We conjecture that Condition α is also necessary for equivalence. Suppose this condition does not hold. Then there exists $P^1, P^{h+1} \in D$ and $x, y \in A$ such that xP^1y and $xP^{h+1}y$ and all paths between P^1 and P^{h+1} contain an essential sequence between y and x. We conjecture that it is possible in this case to construct a social choice function where equivalence breaks down. Note that it is possible to assign alternatives to all preferences on any path between P^1 and P^{h+1} with $f(P^1) = y$ and $f(P^{h+1}) = x$ and satisfying AM-proofness. Strategy-proofness is violated because there exists a manipulation at P^1 via P^{h+1}. If all paths between P^1 and P^{h+1} have no preferences in common except P^1 and P^{h+1}, then the argument is complete. However if the paths have a preference in common, a more refined argument is required. We believe that such a social choice function can be constructed.

4.5 Conclusion

We provide a weaker sufficient condition than Sato (2013b). This condition illustrates the issues involved behind equivalence. We believe that our condition will be helpful in identifying more transparent conditions that guarantee equivalence when rules satisfy multi-agent properties such as unanimity, tops-onlyness, etc.

Bibliography

Amorós, P. (2009): "Eliciting socially optimal rankings from unfair jurors," *Journal of Economic Theory*, 144, 1211–1226.

——— (2010): "Picking the Winners," Tech. rep., Universidad de Málaga, Department of Economic Theory, Málaga Economic Theory Research Center.

——— (2011): "A natural mechanism to choose the deserving winner when the jury is made up of all contestants," *Economics Letters*, 110, 241–244.

——— (2013): "Picking the winners," *International Journal of Game Theory*, 42, 845–865.

——— (2014): "Conditions on the jury for the natural implementation of the deserving winner of a contest," Tech. rep., Universidad de Málaga, Department of Economic Theory, Málaga Economic Theory Research Center.

Amorós, P., L. C. Corchón, and B. Moreno (2002): "The scholarship assignment problem," *Games and Economic behavior*, 38, 1–18.

Archer, A. and R. Kleinberg (2014): "Truthful germs are contagious: a local-to-global characterization of truthfulness," *Games and Economic Behavior*, 86, 340–366.

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

Beviá, C. and L. Corchón (1995): "On the generic impossibility of truthful behavior: A simple approach," *Economic Theory*, 6, 365–371.

Busetto, F. and G. Codognato (2009): "Reconsidering two-agent Nash implementation," *Social Choice and Welfare*, 32, 171–179.

Campbell, D. E. and M. Truchon (1988): "Boundary optima and the theory of public goods supply," *Journal of Public Economics*, 35, 241–249.

Carroll, G. (2012): "When are local incentive constraints sufficient?" *Econometrica*, 80, 661–686.

Corchón, L. C. and J. Rueda-Llano (2008): "Differentiable strategy-proof mechanisms for private and public goods in domains that are not necessarily large or quasi-linear," *Review of Economic Design*, 12, 279–291.

Deb, R. and S. Ohseto (1999): "Strategy-proof and individually rational social choice functions for public good economies: A note," *Economic Theory*, 14, 685–689.

Dogan, B. (2013): "Eliciting the socially optimal allocation from responsible agents," Tech. rep., mimeo, University of Rochester.

Dutta, B. and A. Sen (1991): "A necessary and sufficient condition for two-person Nash implementation," *The Review of Economic Studies*, 58, 121.

——— (2012): "Nash implementation with partially honest individuals," *Games and Economic Behavior*, 74, 154–169.

Gibbard, A. (1977): "Manipulation of schemes that mix voting with chance," *Econometrica: Journal of the Econometric Society*, 665–681.

Goswami, M. P. (2013): "Non fixed-price trading rules in single-crossing classical exchange economies," *Social Choice and Welfare*, forthcoming.

Goswami, M. P., M. Mitra, and A. Sen (2014): "Strategy proofness and Pareto efficiency in quasilinear exchange economies," *Theoretical Economics*, 9, 361–381.

Hashimoto, K. (2008): "Strategy-proofness versus efficiency on the Cobb-Douglas domain of exchange economies," *Social choice and welfare*, 31, 457–473.

Hurwicz, L. (1972): "On Informationally Decentralized Systems." In Decision and Organization: A Volume in Honor of Jacob Marschak, edited by CB McGuire and Roy Radner. Amsterdam: North-Holland," .

Hurwicz, L. and M. Walker (1990): "On the generic nonoptimality of dominant-strategy allocation mechanisms: A general theorem that includes pure exchange economies," *Econometrica: Journal of the Econometric Society*, 683–704.

Jackson, M. O. (2003): "Mechanism Theory," .

Kartik, N. and O. Tercieux (2012): "Implementation with evidence," *Theoretical Economics*, 7, 323–355.

Lombardi, M. (2010): "Two-agent Nash implementation with partially-honest agents: Almost Full Characterizations," *MPRA Paper*.

Lombardi, M. and N. Yoshihara (2011): "Partially-honest Nash implementation: Characterization results," *CCES Discussion Paper Series*.

Matsushima, H. (2008): "Role of honesty in full implementation," *Journal of Economic Theory*, 139, 353–359.

Mishra, D., A. Pramanik, and S. Roy (2015): "Local Incentive Compatibility with Transfers," Working Paper.

Moore, J. and R. Repullo (1990): "Nash Implementation: A Full characterization," *Econometrica*, 58, 1083–1099.

Moulin, H. (1980): "On strategy-proofness and single peakedness," *Public Choice*, 35, 437–455.

Nicoló, A. (2004): "Efficiency and truthfulness with Leontief preferences. A note on two-agent, two-good economies," *Review of Economic Design*, 8, 373–382.

Rudin, W. (1976): *Principles of Mathematical Analysis*, vol. 3, McGraw-Hill New York.

Saporiti, A. (2009): "Strategy-proofness and single-crossing," *Theoretical Economics*, 4, 127–163.

Sato, S. (2013a): "Strategy-proofness and the reluctance to make large lies: the case of weak orders," *Social Choice and Welfare*, 40, 479–494.

——— (2013b): "A sufficient condition for the equivalence of strategy-proofness and non-manipulability by preferences adjacent to the sincere one," *Journal of Economic Theory*, 148, 259–278.

Satterthwaite, M. A. and H. Sonnenschein (1981): "Strategy-proof allocation mechanisms at differentiable points," *The Review of Economic Studies*, 587–597.

Serizawa, S. (1996): "Strategy-proof and individually rational social choice functions for public good economies," *Economic Theory*, 7, 501–512.

——— (2002): "Inefficiency of strategy-proof rules for pure exchange economies," *Journal of Economic Theory*, 106, 219–241.

——— (2006): "Pairwise strategy-proofness and self-enforcing manipulation," *Social Choice and Welfare*, 26, 305–331.

Serizawa, S. and J. A. Weymark (2003): "Efficient strategy-proof exchange and minimum consumption guarantees," *Journal of Economic Theory*, 109, 246–263.

Tian, G. (1996): "On the existence of optimal truth-dominant mechanisms," *Economics Letters*, 53, 17–24.

Weymark, J. A. (2011): "A unified approach to strategy-proofness for single-peaked preferences," *SERIEs*, 2, 529–550.

Zhou, L. (1991): "Inefficiency of strategy-proof allocation mechanisms in pure exchange economies," *Social Choice and Welfare*, 8, 247–254.