Obviously Strategy-Proof Mechanisms[*]

Shengwu Li[†]
Working Paper

First uploaded: 3 Feb 2015. This version: 17 July 2015.

1 Abstract

In mechanism design, strategy-proofness is often said to be desirable because it makes it easy for agents to decide what to do. However, some strategy-proof mechanisms are easier to understand than others. In this paper, I define what it means for a mechanism to be Obviously Strategy Proof (OSP). This is a refinement of strategy-proofness that applies to extensive game forms. Using a formal model of cognitive limitations, I show that a mechanism is OSP iff the optimality of truth-telling can be deduced without contingent reasoning. I show that a choice rule is OSP-implementable iff it can be carried out by a social planner under a particular regime of partial commitment. Finally, I characterize the set of OSP mechanisms in a canonical setting, that encompasses private-value auctions with unit demand and binary public good problems.

2 Introduction

Strategy-proof mechanisms are often said to be desirable. They reduce participation costs and cognitive costs, by making it easy for agents to decide what to do.[1] They protect unsophisticated agents from strategic errors.

---

[1]Vickrey (1961) writes that, in second-price auctions: "Each bidder can confine his efforts and attention to an appraisal of the value the article would have in his own hands, at a considerable saving in mental strain and possibly in out-of-pocket expense."

They prevent waste from rent-seeking espionage. The resulting outcome does not depend sensitively on each agent's beliefs about other agents' preferences or information.

These benefits largely depend on agents *recognizing* that the mechanism is strategy-proof (SP). Only then can they conclude that they need not engage in detailed reasoning about their opponents' strategies.

However, some strategy-proof mechanisms are simpler for real people to understand than others. For instance, ascending clock auctions and second-price sealed-bid auctions are isomorphic under current theories.[2] Nonetheless, laboratory subjects are substantially more likely to play the weakly dominant strategy under a clock auction than under sealed bids. (Kagel et al., 1987) Theorists have also expressed this intuition:

> Some other possible advantages of dynamic auctions over static auctions are difficult to model explicitly within standard economics or game-theory frameworks. For example, ...it is generally held that the English auction is simpler for real-world bidders to understand than the sealed-bid second-price auction, leading the English auction to perform more closely to theory. (Ausubel, 2004)

In this paper, I model explicitly what it means for a mechanism to be *obviously* strategy-proof. This is a refinement of strategy-proofness, that applies to extensive game forms. This approach invokes no new primitives. It identifies a set of mechanisms as obviously strategy-proof, while remaining as parsimonious as the standard theory of extensive games.

A mechanism is *obviously strategy-proof* (OSP) if it has an equilibrium in obviously dominant strategies. A strategy S_i is *obviously dominant* if, for any deviating strategy S_i', starting from any earliest information set where S_i' and S_i disagree, the best possible outcome from S_i' is no better than the worst possible outcome from S_i.

Note that this requirement applies only to earliest information sets where S_i and S_i' disagree. This requirement therefore depends on the extensive game form.

Ascending clock auctions are OSP. Suppose Bidder 1 values the object at $10. In an ascending clock auction, the price rises monotonically. If the current price is below $10, then the best possible outcome from quitting now

---

[2]To be precise, if we restrict our attention to cut-off strategies in ascending clock auctions, then ascending clock auctions and second-price sealed-bid auctions have identical reduced normal form representations.

Figure 1: 'Similar' mechanisms from 1's perspective.

(i)   (ii)   (iii)
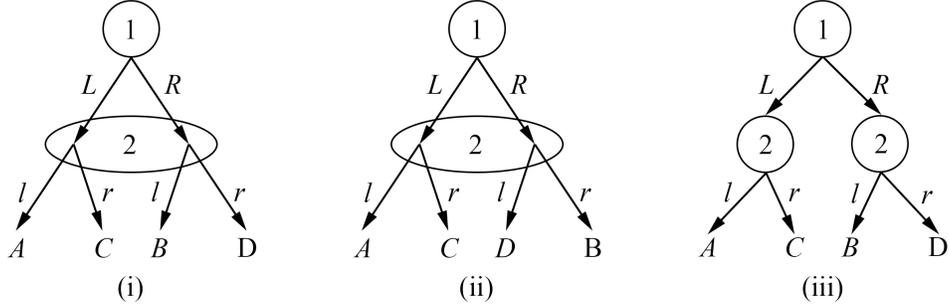
is no better than the worst possible outcome from staying in the auction (and quitting at \$10). If the price is above \$10, then the best possible outcome from staying in the auction is no better than the worst possible outcome from quitting now.

Second-price sealed-bid auctions are SP, but not OSP. Consider the strategies "bid \$10" and "bid \$11". The earliest information set where these disagree is the point where Bidder 1 submits her bid. If Bidder 1 bids \$11, she might win the object at some price strictly below \$10. If Bidder 1 bids \$10, she might not win the object. The best possible outcome from deviating is better than the worst possible outcome from truth-telling. This captures an intuition expressed by experimental economists:

> The idea that bidding modestly in excess of x only increases the chance of winning the auction when you don't want to win is far from obvious from the sealed bid procedure. (Kagel et al., 1987)

I produce two characterization theorems, which suggest two interpretations of OSP.

First, I propose a formal model of a cognitively limited agent. I show that a strategy S_i is obviously dominant if and only if such an agent can recognize S_i as weakly dominant.

Consider the mechanisms in Figure 1. Suppose Agent 1 has preferences: $A \succ B \succ C \succ D$. In (i), it is a weakly dominant strategy for 1 to play L. All three mechanisms are intuitively similar, but it is not a weakly dominant strategy for Agent 1 to play L in (ii) and (iii).

In order for Agent 1 to recognize that it is weakly dominant to play L in mechanism (i), he must think through hypothetical scenarios case-by-case:

3

"If Agent 2 plays l, then I should play L, since I prefer A to B. If Agent 2 plays r, then I should play L, since I prefer C to D. Therefore, I should play L, no matter what agent 2 plays." This mental process is called *contingent reasoning*. Notice that the quoted inferences are valid in (i), but not valid in (ii) and (iii).

Suppose Agent 1 is unable to engage in contingent reasoning. That is, he knows that playing L might lead to A or C, and playing R might lead to B or D. However, he does not understand how, case-by-case, the outcomes after playing L are related to the outcomes after playing R. Then it is as though he cannot distinguish (i), (ii), and (iii).

This idea can be made formal and general. I define an equivalence relation on the space of mechanisms: The *experience* of agent i at history h records the information sets where i was called to play, and the actions that i took, in chronological order.[3] Two mechanisms G and G' *generate the same experiences* for i if there is a bijection from i's information sets and actions in G, onto i's information sets and actions in G', such that:

1. G can produce for i some experience if and only if G' can produce for i its bijected partner experience.

2. An experience might result in some outcome in G if and only if its bijected partner might result in that same outcome in G'.

With this relation, we can partition the set of all mechanisms into equivalence classes. For instance, all three mechanisms in Figure 1 generate the same experiences for Agent 1.

Within a mechanism, a partition of histories into information sets represents imperfect information about past actions. Similarly, a partition of mechanisms into equivalence classes represents imperfect understanding about the properties of each mechanism. Any such partition is, implicitly, a model of a cognitively limited agent.

The partition defined by the relation "G and G' generate the same experiences for i" rules out contingent reasoning. Suppose an agent is unable to distinguish games that generate the same experiences. He retains substantial knowledge about the structure of the game. He knows the information sets at which he may be called to play, and the actions available at each information set. He knows, for any experience, what outcomes may result. However, he is unable to reason case-by-case about hypothetical scenarios.

---

[3]An experience is a standard concept in the theory of extensive games; experiences are used to define perfect recall.

The first characterization theorem states: A strategy S_i is obviously dominant in G if and only if it is weakly dominant in every G' that generates the same experiences for i as G.

This shows that obviously dominant strategies are those that can be recognized as weakly dominant without contingent reasoning. An obviously dominant strategy is weakly dominant in any game that generates the same experiences. In that sense, such a strategy is robustly dominant.

The Prisoner's Dilemma is a special case of game (i) in Figure 1; playing defect is not obviously dominant. On the other hand, if Agent 1 is informed of Agent 2's action before making his decision, then playing defect is obviously dominant. Shafir and Tversky (1992) find that laboratory subjects in a Prisoner's Dilemma are more likely to play the weakly dominant strategy when they are informed beforehand that their opponent has cooperated (84%) or when they are informed beforehand that their opponent has defected (97%), compared to when they are not informed of their opponent's strategy (63%).[4] This violates the Sure Thing Principle (Savage, 1954), but is predicted by the cognitive model that I have just exposited.

The second characterization theorem for OSP relates to the problem of mechanism design under partial commitment. In mechanism design, we usually assume that the Planner can commit to every detail of a mechanism, including the events that an individual agent does not directly observe. For instance, in a sealed-bid auction, it is assumed that the Planner can commit to the function from all sealed bids to allocations and payments, even though each agent only directly observes his own bid. In some settings, this is unrealistic. If agents cannot individually verify the details of a mechanism, the Planner may be unable to commit to it.

Mechanism design under partial commitment is a pressing problem. Auctions run by central brokers over the Internet account for billions of dollars of economic activity. (Edelman et al., 2007) In such settings, bidders may be unable to verify that the other bidders exist, let alone what actions they have taken. As another example, some wireless spectrum auctions use computationally demanding techniques to solve complex assignment problems.

---

[4]82% of subjects defect *both* when informed that their opponent has cooperated, and when informed that their opponent has defected. Note that *random* choosers would play the weakly dominant strategy 50% of the time. In Shafir and Tversky's data, switching from coin flips to human beings with weakly dominant strategies yields a 13% increase in the rate that the dominant strategy is played. Switching from weakly dominant strategies to obviously dominant strategies yields a further increase of 19%. This suggests that, in addition to social preferences, cognitive limitations have a role to play in explaining the data.

In these settings, individual bidders may find it difficult and costly to verify the output of the auctioneer's algorithm. (Milgrom and Segal, 2014)

For the second characterization theorem, I consider a 'metagame' where the Planner privately communicates with agents, and eventually decides on an outcome. The Planner chooses one agent, and sends a private message, along with a set of acceptable replies. That agent chooses a reply, which the Planner observes. The Planner can then either repeat this process (possibly with a different agent) or announce an outcome and end the game.

The Planner has partial commitment power: For each agent, she can commit to use only a subset of her available strategies. However, the subset she offers to agent i must be measurable with respect to i's observations in the game. We call this a *bilateral commitment*. Suppose, for robustness, that each agent could hold *any* beliefs about the Planner's strategy, subject to the constraints of his bilateral commitment. What choice rules can be implemented in this metagame?

The second characterization theorem states: A choice rule can be supported by bilateral commitments if and only if that choice rule is OSP-implementable.

Thus, in addition to formalizing a notion of cognitive simplicity, OSP also captures the set of choice rules that can be carried out with only bilateral commitments.

After defining and characterizing OSP, I apply this new concept to the environment of binary allocation problems. In this environment, there is a set of agents N with continuous single-dimensional types $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i]$. An *allocation* y is a subset of N. An *allocation rule* f_y is a function from type profiles to allocations. We augment this with a *transfer rule* f_t, which specifies money transfers for each agent. Each agent has utility equal to his type if he is in the allocation, plus his net transfer.

$$u_i(\theta_i, y, t) = 1_{i \in y}\theta_i + t_i \tag{1}$$

Binary allocation problems encompass several canonical settings. They include private-value auctions with unit demand. They include procurement auctions with unit supply; not being in the allocation is 'winning the contract', and the bidder's type is his cost of provision. They also include binary public good problems; the feasible allocations are N and the empty set.

Mechanism design theory has extensively investigated SP-implementation in this environment. f_y is SP-implementable if and only if f_y is monotone in each agent's type (i.e. $1_{i \in f_y(\theta)}$ is weakly increasing in θ_i). (Spence, 1974;

Mirrlees, 1971; Myerson, 1981) If f_y is SP-implementable, then the accompanying transfer rule f_t is essentially unique. (Green and Laffont, 1977; Holmström, 1979)

$$f_{t,i}(\theta_i, \theta_{-i}) = -1_{f_y(\theta) \in Y_i} \inf\{\theta_i' : i \in f_y(\theta_i', \theta_{-i})\} + r_i(\theta_{-i}) \qquad (2)$$

where r_i is some arbitrary deterministic function of the other agents' preferences.

What are analogues of these canonical results, if we require OSP-implementation rather than SP-implementation? Are ascending clock auctions special, or are there other OSP mechanisms in this setting?

I develop an analogous essential uniqueness theorem for extensive game forms: Every mechanism that OSP-implements a private-value allocation rule is 'essentially' a monotone price mechanism, which is a generalization of ascending clock auctions.

This theorem implies that when we desire OSP-implementation in this environment, we need not search the (very large) space of all extensive game forms. We can focus our attention on the class of monotone price mechanisms.

Next, I characterize the set of OSP-implementable allocation rules. For this part, I assume that the lowest type of each agent is never in the allocation, and is required to have a zero transfer.[5] Given an allocation rule, I show how to identify subsets of $\mathbb{R}^{|N|}$ that contain viable price paths for a monotone price mechanism. I provide a necessary and sufficient condition for an allocation rule to be OSP-implementable.

As a second application, I consider a generalization of the Edelman et al. (2007) online advertising environment. In this setting, agents bid for advertising positions, each worth a certain number of clicks. Each agent's type is a vector of per-click values, one for each position. I show that if preferences satisfy a single-crossing condition, then we can OSP-implement the efficient allocation and the Vickrey payments.

As a third application, I produce an impossibility result for a classic matching algorithm: With 3 or more agents, there does not exist a mechanism that OSP-implements Top Trading Cycles. (Shapley and Scarf, 1974)

The rest of the paper proceeds in the usual order. Proofs omitted from the main text are in Appendix C.

---

[5]This assumption is borrowed from the Revenue Equivalence Theorem. (Myerson, 1981)

3   Related Literature

It is generally acknowledged that ascending auctions are simpler for real bidders than sealed-bid auctions. (Ausubel, 2004) Laboratory experiments have investigated and corroborated this claim (Kagel et al., 1987; Kagel and Levin, 1993). More generally, Charness and Levin (2009) and Esponda and Vespa (2014) document that laboratory subjects find it difficult to reason case-by-case about hypothetical scenarios. This mental process is often called "contingent reasoning", but has received little formal treatment in economic theory.[6]

There is also a strand of literature, including Vickrey's seminal paper, that observes that sealed-bid auctions raise problems of commitment. (Vickrey, 1961; Rothkopf et al., 1990; Cramton, 1998) For instance, it may be difficult to prevent shill bidding without third-party verification. Rothkopf et al. (1990) argue that "robustness in the face of cheating and of fear of cheating is important in determining auction form".

This paper formalizes and unifies both these strands of thought. It shows that mechanisms that do not require contingent reasoning are identical to mechanisms that can be run under bilateral commitment.

This paper also relates to the planned US auction to repurchase television broadcast rights. In this setting, complex underlying constraints have the result that Vickrey prices cannot be computed without large approximation errors. Milgrom and Segal (2014) propose the use of a clock auction to repurchase broadcast rights. They recommend this over an equivalent sealed-bid procedure, arguing that clock auctions "make strategy-proofness self-evident even for bidders who misunderstand or mistrust the auctioneer's calculations". The Milgrom-Segal clock auction uses advanced computational techniques to solve a challenging allocation problem. However, it is obviously strategy-proof.

In multi-unit combinatorial auction problems, finding the optimal solution is NP-hard, so the Vickrey-Clarke-Groves mechanism may be computationally infeasible. Bartal et al. (2003) propose an approximating mechanism that runs in polynomial time. In this mechanism, bidders are processed sequentially. Each bidder is presented with a price vector and chooses his most-preferred bundle, and his choice is used to adjust the price vectors

---

[6]There is a theory literature that studies agents that do not fully account for other agents' private information. (Eyster and Rabin, 2005; Crawford and Iriberri, 2007; Esponda, 2008) This is a related notion, but conceptually distinct from mistakes in contingent reasoning. In particular, these models predict no deviations from classical play in strategy-proof mechanisms.

for the remaining bidders. One previously unmodeled advantage of such mechanisms is that they are obviously strategy-proof.

OSP requires equilibrium in obviously dominant strategies. This is distinct from O-solvability, a solution concept used in the computer science literature on decentralized learning. (Friedman, 2002, 2004) Strategy S_i is said to *overwhelm* S_i' if the worst possible outcome from S_i is strictly better than the best possible outcome from S_i'. O-solvability calls for the iterated deletion of overwhelmed strategies. One difference between the two concepts is that O-solvability is for normal form games, whereas OSP invokes a notion of an 'earliest point of departure', which is only defined in the extensive form.[7]

4   Definition and Characterization

The planner operates in an *environment* consisting of:

1. A set of agents, $N \equiv \{1, \ldots, n\}$.

2. A set of outcomes, X.

3. A set of preference profiles over X, $\Theta \equiv \prod_{i \in N} \Theta_i$.

4. A set of game forms with consequences in X, $\mathcal{G}$.

An *extensive game form with consequences in X* is a tuple $\langle N, H, P, \delta_c, (\mathcal{I}_i)_{i \in N}, g \rangle$, where:

1. N is a set of agents.

2. H is a set of histories (sequences of actions, including the null sequence $h_\emptyset$).

   (a) Let Z be the terminal histories. $z \in Z$ iff there does not exist $h \in H$ such that z is a proper subhistory of h.[8]

   (b) Let $A(h)$ be the actions available at h.

3. P is a player function. $P : H \setminus Z \to N \cup c$

---

[7]Ascending clock auctions and sequential random serial dictatorship are OSP, but not O-solvable. O-solvability is too strong for our current purpose.

[8]As usual, this implies that infinitely long histories are terminal histories.

4. δ_c is the chance function. For any h where $P(h) = c$, δ_c specifies a probability measure on $A(h)$.[9]

5. $\mathcal{I}_i$ is a partition of $\{h : P(h) = i\}$ with the property that $A(h) = A(h')$ whenever h and h' are in the same member of the partition.

   (a) For any $I_i \in \mathcal{I}_i$, we denote: $P(I_i) = P(h)$ for any $h \in I_i$. $A(I_i) = A(h)$ for any $h \in I_i$.

   (b) If (a, a') are such that $a \in A(I_i)$, $a' \in A(I'_j)$, $I_i \neq I'_j$ then we regard a and a' as distinct.

6. g is an outcome function. It associates each terminal history with an outcome. $g : Z \to X$

A *strategy* S_i for agent i in game G specifies what agent i does at every one of her information sets. $S_i(I_i) \in A(I_i)$. A *strategy profile* $S = (S_i)_{i \in N}$ is a set of strategies, one for each agent. When we want to refer to the strategies used by different types of i, we use $S_i^{\theta_i}$ to denote the strategy assigned to type θ_i.

Let $z^G(h, S, \delta_c)$ be the lottery over terminal histories that results in game form G when we start from h and play proceeds according to (S, δ_c). $z^G(h, S, d_c)$ is the result of one realization of the chance moves under δ_c. We sometimes write this as $z^G(h, S_i, S_{-i}, d_c)$.

$u_i(x, \theta_i) \in \mathbb{R}$ is the utility to agent i from outcome x given preferences θ_i. Let $u_i^G(h, S_i, S_{-i}, d_c, \theta_i) \equiv u_i(g(z^G(h, S_i, S_{-i}, d_c)), \theta_i)$. This is the utility to agent i in game G, when we start at history h, play proceeds according to (S_i, S_{-i}, d_c), and the resulting outcome is evaluated according to preferences θ_i.

We require that δ_c has full support on the available moves $A(h)$ when it is called to play.[10] This is without loss of generality, since if some action is not in the support of the chance function, then we can simply delete that action from the game.

Definition 1. *$\psi_i(h)$ is the experience of agent i along history h. $\psi_i(h)$ is an alternating sequence of information sets and actions. It is constructed as follows: Let $h_1, \ldots, h_K$ be the subhistories of h where $P(h) = i$, in order.*

---

[9]We require the probability distributions to be independent across different histories.

[10]This ensures that obvious dominance has a useful invariance property: Namely, that the set of obviously dominant strategies does not change when we add new chance moves that occur with zero probability. Appendix A defines full support in a way that accommodates infinite and uncountably infinite action sets.

1. *The $(2k-1)$th element in $\psi_i(h)$ is $I_i : h_k \in I_i$.*

2. *If h_k is a proper subhistory of h, then the $(2k)$th element in $\psi_i(h)$ is the action that immediately follows h_k in h.*

We use Ψ_i to denote the set $\{\psi_i(h) : h \in H \text{ and } \psi_i(h) \neq \{\emptyset\}\}$.

An extensive game form has *perfect recall* if for any information set I_i, for any two histories h and h' in I_i, $\psi_i(h) = \psi_i(h')$. We use $\psi_i(I_i)$ to denote $\psi_i(h) : h \in I_i$.

Definition 2. *$\mathcal{G}$ is the set of all extensive game forms with consequences in X and perfect recall, where δ_c has full support.*

A *choice rule* is a function $f : \Theta \to X$. If we consider stochastic choice rules, then it is a function $f : \Theta \to \Delta X$.[11]

A *solution concept* $\mathcal{S}$ is a set-valued function with domain $\mathcal{G} \times \Theta$. It takes values in the set of strategy profiles.

Definition 3. *f is $\mathcal{S}$-implementable if there exists*

1. *$G \in \mathcal{G}$*

2. *$(S^\theta)_{\theta \in \Theta} \equiv ((S_i^{\theta_i})_{i \in N})_{\theta \in \Theta}$*

such that, for all $\theta \in \Theta$

1. *$S^\theta \in \mathcal{S}(G, \theta)$.*

2. *$f(\theta) = g(z^G(\emptyset, S^\theta, \delta_c))$*

Note that our concern is with weak implementation: We require that $S^\theta \in \mathcal{S}(G, \theta)$, not $\{S^\theta\} = \mathcal{S}(G, \theta)$. This is to preserve the analogy with canonical results for strategy-proofness, many of which assume weak implementation. (Myerson, 1981; Saks and Yu, 2005)

We use "$(G, (S^\theta)_{\theta \in \Theta})$ $\mathcal{S}$-implements f" to mean that $(G, (S^\theta)_{\theta \in \Theta})$ fulfils the requirements of Definition 3. We use "G $\mathcal{S}$-implements f" to mean that there exists $(S^\theta)_{\theta \in \Theta}$ such that $(G, (S^\theta)_{\theta \in \Theta})$ fulfils the requirements of Definition 3.

---

[11] For readability, we generally suppress the latter notation, but the claims that follow hold for both deterministic and stochastic choice rules. Additionally, the set X could *itself* be a set of lotteries. The interpretation of this is that the planner can carry out one-time public lotteries at the end of the mechanism, where the randomization is observable and verifiable.

Definition 4 (Weakly Dominant). *In G for agent i with preferences θ_i, S_i is **weakly dominant** if:*

$$\forall S_i' : \forall S_{-i}' :$$
$$\mathbb{E}_{\delta_c}[u_i^G(h_\emptyset, S_i', S_{-i}', d_c, \theta_i)] \leq \mathbb{E}_{\delta_c}[u_i^G(h_\emptyset, S_i, S_{-i}', d_c, \theta_i)] \tag{3}$$

Let $\alpha(S_i, S_i')$ be the set of earliest points of departure for S_i and S_i'. That is, $\alpha(S_i, S_i')$ contains the information sets where S_i and S_i' have made identical decisions at all prior information sets, but are making a different decision now. We define this recursively as follows:[12]

Definition 5 (Earliest Points of Departure). *$I_i \in \alpha(S_i, S_i')$ if and only if:*

1. *$S_i(I_i) \neq S_i'(I_i)$*

2. *There exists $h \in I_i, S_{-i}, d_c$ such that h is a subhistory of $z^G(h_\emptyset, S_i, S_{-i}, d_c)$.[13]*

3. *There does not exist (h, h', I_i') such that:*

   (a) *h' is a proper subhistory of h*

   (b) *$h \in I_i$*

   (c) *$h' \in I_i'$*

   (d) *$I_i' \in \alpha(S_i, S_i')$*

Definition 6 (Obviously Dominant). *In G for agent i with preferences θ_i, S_i is **obviously dominant** if:*

$$\forall S_i' : \forall I_i \in \alpha(S_i, S_i') :$$
$$\sup_{h \in I_i, S_{-i}', d_c} u_i^G(h, S_i', S_{-i}', d_c, \theta_i) \leq \inf_{h \in I_i, S_{-i}', d_c} u_i^G(h, S_i, S_{-i}', d_c, \theta_i) \tag{4}$$

---

[12]The perfect recall assumption is necessary for this to work. Suppose a game of imperfect recall, with one agent, with histories $\{\emptyset, L, R, RL, RR\}$, with information set $I_1 = \{\emptyset, R\}$. Then the strategies $S_1(I_1) = L$ and $S_1(I_1) = R$ have no earliest point of departure. This recalls Piccione and Rubinstein's paradox of the absentminded driver. (Piccione and Rubinstein, 1997)

[13]This is just saying that S_i does not itself rule out I_i from the path of play. Even though this requirement is not stated symmetrically, for all (S_i, S_i'), $\alpha(S_i, S_i') = \alpha(S_i', S_i)$.

Compare Definition 4 and Definition 6. Weak dominance is defined using $h_\emptyset$, the history that begins the game. Consequently, if two extensive games have the same normal form, then they have the same weakly dominant strategies. Obvious dominance is defined with histories that are in information sets that are earliest points of departure. Thus two extensive games with the same normal form may not have the same obviously dominant strategies. Switching to a direct revelation mechanism may not preserve obvious dominance, so the standard revelation principle does not apply.

In what sense is obvious dominance obvious? Definition 4 contains the quantifier "$\forall S_{-i}$". To assess that S_i weakly dominates S_i', one has to keep track of as many inequalities as there are combinations of opponent strategies. By contrast, Definition 6 does not contain "$\forall S_{-i}$". To assess that S_i obviously dominates S_i' requires only one inequality at each earliest point of departure. Thus, for agents who find case-by-case reasoning difficult, obvious dominance may be easier to assess than weak dominance.

Definition 7 (Strategy-Proof). $S \in \mathbf{SP}(G, \theta)$ *if for all i, S_i is weakly dominant.*

Definition 8 (Obviously Strategy-Proof). $S \in \mathbf{OSP}(G, \theta)$ *if for all i, S_i is obviously dominant.*

Consider the case where there is a set of $|N|$ objects and each agent's type is a vector of per-object values. *Sequential Random Serial Dictatorship* refers to the procedure where, in a random order, each agent takes one object from the set that remains.

Proposition 1. *Sequential Random Serial Dictatorship is obviously strategy-proof.*

Proof. By inspection. □

Proposition 2. *If G is obviously strategy-proof, then G is weakly group-strategy-proof.*

Proof. Take any type profile θ. Take $S \in \mathbf{OSP}(G, \theta)$. Suppose there was a coalition $\hat{N} \subseteq N$ that could jointly deviate to strategies $(\hat{S}_i)_{i \in \hat{N}}$ and all be strictly better off. Fix $(S_i)_{i \in (N \setminus \hat{N})}$ and d_c such that all agents in the coalition are strictly better off. Along the resulting terminal history, there must be a first agent i in the coalition to deviate from S_i to $\hat{S}_i$. That first deviation happens at some information set $I_i \in \alpha(S_i, \hat{S}_i)$. Since $S \in \mathbf{OSP}(G, \theta)$, agent i cannot strictly gain from deviating to $\hat{S}_i$; a contradiction. □

Corollary 1. *If G is obviously strategy-proof, then G is strategy-proof.*

4.1 Cognitive limitations

We now define an equivalence relation between mechanisms. In words, G and G' generate the same experiences for i if there exists a bijection from i's information sets and actions in G onto i's information sets and actions in G', such that:

1. ψ_i is an experience in G iff ψ_i's bijected partner is an experience in G'.

2. Outcome x could follow experience ψ_i in G iff x could follow ψ_i's bijected partner in G'

Definition 9. *Take any $G, G' \in \mathcal{G}$, with information partitions $\mathcal{I}_i, \mathcal{I}_i'$ and experience sets Ψ_i, Ψ_i'. G and G' **generate the same experiences for** i if there exists a bijection λ from $\mathcal{I}_i \cup A(\mathcal{I}_i)$ onto[14] $\mathcal{I}_i' \cup A'(\mathcal{I}_i')$ such that:*

1. $\psi_i \in \Psi_i$ iff $\lambda(\psi_i) \in \Psi_i'$

2. $\exists z \in Z : g(z) = x, \psi_i(z) = \psi_i$ iff $\exists z' \in Z' : g'(z') = x, \psi_i'(z') = \lambda(\psi_i)$

where we use $\lambda(\psi_i)$ to denote $\{\lambda(\psi_i^k)\}_{k=1}^T$, where $T \in \mathbb{N} \cup \infty$.

For G and G' that generate the same experiences, we define $\lambda(S_i)$ to be the strategy that, given information set I_i' in G', plays $\lambda(S_i(\lambda^{-1}(I_i')))$.

Theorem 1. *For any i, θ_i: Consider the equivalence classes on $\mathcal{G}$ defined by the relation "G and G' generate the same experiences for i". S_i is obviously dominant in G if and only if for every G' in the equivalence class of G, $\lambda(S_i)$ is weakly dominant in G'.*

The "if" direction permits a constructive proof. Suppose S_i is not obviously dominant in G. Then we can take a sequence of 'experience-preserving' transformations of G, to produce some G' where $\lambda(S_i)$ is not weakly dominant. The "only if" direction proceeds as follows: Suppose there exists some G' in the equivalence class of G, where $\lambda(S_i)$ is not weakly dominant. Then we can use λ^{-1} to locate an information set in G, and a deviation S_i', that do not satisfy the obvious dominance inequality. Appendix C provides the details.

---

[14]This definition entails that λ maps $\mathcal{I}_i$ onto $\mathcal{I}_i'$ and $A(\mathcal{I}_i)$ onto $A'(\mathcal{I}_i')$. If an information set in G was mapped onto an action in G', then any experience involving that information set would, when passed through the bijection, result in a sequence that was not an experience, and *ipso facto* not an experience of G'.

4.2 Supported by bilateral commitments

Suppose the following game: As before we have a set of agents N, outcomes X, and preference profiles $\prod_{i \in N} \Theta_i$. However, there is one player in addition to N: Player 0, the Planner.

The Planner has an uncountably infinite message space M. At the start of the game, each agent $i \in N$ privately observes θ_i. Play proceeds as follows:

1. The Planner chooses one agent $i \in N$ and sends a query $m \in M$, along with a set of acceptable replies $R \subset M$

2. i observes (m, R), and chooses a reply $r \in R$.

3. The Planner observes r.

4. The Planner either selects an outcome $x \in X$, or chooses to send another query.

   (a) If the Planner selects an outcome, the game ends.

   (b) If the Planner chooses to send another query, go to Step 1.

We allow the Planner's strategy $\tilde{S}_0$ to assign outcomes to infinitely long sequences of queries and replies.

For $i \in N$, i's strategy specifies what reply to give, as a function of her preferences, the past sequence of queries and replies between her and the Planner, and the current (m, R). That is

$$\tilde{S}_i(\theta_i, (m_k, R_k, r_k)_{k=1}^{t-1}, m_t, R_t) \in R_t \tag{5}$$

We use $\tilde{S}_i^{\theta_i}$ to denote the strategy played by type θ_i of agent i.

Let Ψ_0 be the set of all Planner strategies. The standard full commitment paradigm is equivalent to allowing the Planner to commit to a unique $\tilde{S}_0 \in \Psi_0$. Instead, we assume that for each agent, the Planner can commit to a subset $\hat{\Psi}_0^i \subseteq \Psi_0$ that is measurable with respect to that agent's observations in the game.

γ_i is some full sequence of messages and responses seen by player i, $(m_k, R_k, r_k)_{k=1}^T$ for $T \in \mathbb{N} \cup \infty$, along with the resulting outcome. Γ_i denotes a *communication event*, which is some set of such objects. Let $\mu(\Gamma_i, \tilde{S}_0, \tilde{S}_N)$ be the probability that some $\gamma_i \in \Gamma_i$ occurs when the strategies are $(\tilde{S}_0, \tilde{S}_N)$.

Definition 10. $\hat{\Psi}_0^i$ *is a **bilateral commitment** if there exists Γ_i such that: $\tilde{S}_0 \in \hat{\Psi}_0^i$ if and only if*

$$\forall \tilde{S}_N : \mu(\Gamma_i, \tilde{S}_0, \tilde{S}_N) = 0 \tag{6}$$

A bilateral commitment is (equivalently) specified by a 'prohibited' set of observations Γ_i for player i. Bilateral commitments are those such that, if the planner reneges, then agent i might detect this reneging (i.e. observe some $\gamma_i \in \Gamma_i$) with positive probability.

Definition 11. *A choice rule f is **supported by bilateral commitments** $(\hat{\Psi}_0^i)_{i \in N}$, if there exists $(\tilde{S}_0, \tilde{S}_N)$ such that:*

1. *For all θ: $(\tilde{S}_0, \tilde{S}_N)$ results in $f(\theta)$.*

2. *For all $i \in N$: $\tilde{S}_0 \in \hat{\Psi}_0^i$*

3. *For all $i \in N$, for all θ_i, for all $\tilde{S}'_{N \setminus i}$, for all $\tilde{S}'_0 \in \hat{\Psi}_0^i$: $\tilde{S}_i^{\theta_i}$ is a best response to $\tilde{S}'_{N \setminus i}$ and $\tilde{S}'_0$*

The first requirement states that the strategy profile must result in the outcome prescribed by the choice rule. The second requirement is that the Planner's strategy is compatible with the commitment she offers to each agent. The third requirement is that each agent i's strategy $\tilde{S}_i$ is weakly dominant, when we consider the Planner as just another player whose strategies are confined to $\hat{\Psi}_0^i$.

"Supported by bilateral commitments" is just one of many partial commitment regimes. This one requires that the commitment offered to each agent is measurable with respect to events that she can observe. In reality, contracts are seldom enforceable unless each party can observe breaches. Thus, "supported by bilateral commitments" is a natural case to study.

Theorem 2. *f is OSP-implementable if and only if there exist bilateral commitments $(\hat{\Psi}_0^i)_{i \in N}$ that support f.*

The intuition behind the proof is as follows: A bilateral commitment $\hat{\Psi}_0^i$ is *almost* equivalent to the Planner committing to 'run' only games in some equivalence class of $\mathcal{G}$, for the equivalence classes in Theorem 1. To be precise, a bilateral commitment is equivalent to committing to run mechanisms in the closure of such an equivalence class, where the distance function is with respect to the total variation distance of chance moves. In defining $\mathcal{G}$, we assumed that δ_c has full support on the available actions when it is called to play. A bilateral commitment permits the Planner to run a mechanism where some chance actions are never played. Such a mechanism

does not generate the same experiences as the full-support versions, but we can get arbitrarily close to it without leaving the equivalence class.[15]

Consequently, we can find a set of bilateral commitments that support f_y iff we can find some $(G, (S^\theta)_{\theta \in \Theta})$ such that, for every i, for every θ_i, for every G' that generates the same experiences for i, $\lambda(S_i^{\theta_i})$ is weakly dominant in G'. By Theorem 1, this holds iff f_y is OSP-implementable. Appendix C provides the details.

5 Applications

5.1 Binary Allocation Problems

We now consider a canonical environment, $(N, X, \Theta, \mathcal{G})$. N and $\mathcal{G}$ remain general as before. Let $Y \subseteq 2^N$ be the set of feasible *allocations*, with representative element $y \in Y$. An outcome consists of an allocation $y \in Y$ and a transfer for each agent, $X = Y \times \mathbb{R}^n$. $t \equiv (t_i)_{i \in N}$ denotes a profile of transfers.

Preferences are quasilinear. $\Theta = \prod_{i \in N} \Theta_i$, where $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i]$, for $0 \leq \underline{\theta}_i < \overline{\theta}_i < \infty$. For $\theta_i \in \Theta_i$

$$u_i(\theta_i, y, t) = 1_{i \in y}\theta_i + t_i \tag{7}$$

For instance, in a private value auction with unit demand, $i \in y$ iff agent i receives at least one unit of the good under allocation y. In a procurement auction, $i \in y$ iff i does not incur costs of provision under allocation y. θ_i is agent i's cost of provision (equivalently, benefit of non-provision). In a public goods game, $Y = \{\emptyset, N\}$.

An *allocation rule* is $f_y : \Theta \to Y$. A choice rule is thus a combination of an allocation rule and a payment rule, $f = (f_y, f_t)$, where $f_t : \Theta \to \mathbb{R}^n$. Similarly, for each game form G, we disaggregate the outcome function, $g = (g_y, g_t)$. In this part, we concern ourselves only with deterministic allocation rules and payment rules, and thus suppress notation involving δ_c and d_c.

Definition 12. *An allocation rule f_y is $\mathcal{S}$-implementable if there exists f_t such that (f_y, f_t) is $\mathcal{S}$-implementable.*

Definition 13. *f_y is monotone if for all i, for all θ_{-i}, $1_{i \in f_y(\theta)}$ is weakly increasing in θ_i.*

---

[15]One tempting modification of the theory is to define equivalence classes to include their closures with respect to chance moves. This would result in a non-transitive equivalence relation.

In this environment, f_y is SP-implementable if and only if f_y is monotone. This result is implicit in Spence (1974) and Mirrlees (1971), and is proved explicitly in Myerson (1981).[16]

Moreover, if an allocation rule f_y is SP-implementable, then the accompanying payment rule f_t is essentially unique.

$$f_{t,i}(\theta_i, \theta_{-i}) = -1_{f_y(\theta) \in Y_i} \inf\{\theta_i' : i \in f_y(\theta_i', \theta_{-i})\} + r_i(\theta_{-i}) \qquad (8)$$

where r_i is some arbitrary deterministic function of the other agents' preferences. This follows easily by arguments similar to those in Green and Laffont (1977) and Holmström (1979).

We are interested in how these results change when we require OSP-implementation. In particular:

1. What condition on f_y characterizes the set of OSP-implementable allocation rules?

2. For OSP-implementation, is there an analogous 'essential uniqueness' result on the extensive game form G?

Definition 14 (Pruning)**.** *Take any $G = \langle N, H, P, (\mathcal{I}_i)_{i \in N}, g \rangle$, and $(S^\theta)_{\theta \in \Theta}$. $\mathcal{P}(G, (S^\theta)_{\theta \in \Theta}) \equiv \langle N, \tilde{H}, \tilde{P}, (\tilde{\mathcal{I}}_i)_{i \in N}, \tilde{g} \rangle$ is the **pruning** of $(G, (S^\theta)_{\theta \in \Theta})$, constructed as follows:*

1. *$\tilde{H} = \{h \in H : \exists \theta : h \text{ is a subhistory of } z^G(\emptyset, S^\theta)\}$*

2. *For all i, if $I_i \in \mathcal{I}_i$ then $(I_i \cap \tilde{H}) \in \tilde{\mathcal{I}}_i$.[17]*

3. *$(\tilde{P}, \tilde{g})$ are (P, g) restricted to domain $\tilde{H}$.*

We now define a monotone price mechanism. Informally, a monotone price mechanism is such that, for every i,

1. Either

   (a) We present i with a fixed transfer associated with *not* being in the allocation.

---

[16] These monotonicity results for are for weak SP-implementation rather than full SP-implementation implementation. Weak SP-implementation requires $S^\theta \in \text{SP}(G, \theta)$. Full SP-implementation requires $\{S^\theta\} = \text{SP}(G, \theta)$. There are monotone allocation rules for which the latter requirement cannot be satisfied. For example, suppose two agents with unit demand. Agent 1 receives one unit iff $v_1 > .5$. Agent 2 receives one unit iff $v_2 > v_1$.

[17] Note that the empty history $h_\emptyset$ is distinct from the empty set. That is to say, $(I_i \cap \tilde{H}) = \emptyset$ does *not* entail that $\{h_\emptyset\} \in \tilde{\mathcal{I}}_i$.

(b) And a going transfer associated with being in the allocation.

2. Or

    (a) We present i with a fixed transfer associated with being in the allocation.

    (b) And a going transfer associated with *not* being in the allocation.

3. The going transfer falls monotonically. (Equivalently, the going price rises.)

4. Whenever the going transfer strictly falls, i has the option to quit, taking the fixed transfer.

5. If the going transfer could fall in future, i has a unique non-quitting action.

The "either" clause contains ascending clock auctions as a special case. The "or" clause contains descending price procurement auctions; agents that do not win the contract receive a fixed zero transfer. There is a positive payment associated with winning the contract (i.e. *not* being in the allocation), which starts at a high level and counts downwards.

Definition 15 (Monotone Price Mechanism)**.** *A game G is a monotone price mechanism if, for every $i \in N$, at every earliest information set I_i^* such that $A(I_i^*) > 1$:*

1. *Either: There exists a real number t_i^0, a function $\tilde{t}_i^1 : \{I_i : I_i^* \in \psi_i(I_i)\} \to \mathbb{R}$, and a set of actions A^0 such that:*

    (a) For all $a \in A^0$, for all z such that $a \in \psi_i(z)$: $i \notin g_y(z)$ and $g_{t,i}(z) = t_i^0$.

    (b) $A^0 \cap A(I_i^*) \neq \emptyset$.

    (c) For all $I_i', I_i'' \in \{I_i : I_i^* \in \psi_i(I_i)\}$.

        i. If $I_i' \in \psi_i(I_i'')$, then $\tilde{t}_i^1(I_i') \geq \tilde{t}_i^1(I_i'')$.

        ii. If I_i' is the penultimate information set in $\psi_i(I_i'')$ and $\tilde{t}_i^1(I_i') > \tilde{t}_i^1(I_i'')$, then $A^0 \cap A(I_i'') \neq \emptyset$.

        iii. If $I_i' \in \psi_i(I_i'')$ and $\tilde{t}_i^1(I_i') > \tilde{t}_i^1(I_i'')$, then $|A(I_i') \setminus A^0| = 1$.

    (d) For all z where $I_i^* \in \psi_i(z)$:

        i. Either: $i \notin g_y(z)$ and $g_{t,i}(z) = t_i^0$.

19

    ii. Or: $i \in g_y(z)$ and $g_{t,i}(z) = \tilde{t}_i^1(I_i')$, where I_i' is the last information set in $\psi_i(z)$.

2. *Or: There exists a real number t_i^1, a function $\tilde{t}_i^0 : \{I_i : I_i^* \in \psi_i(I_i)\} \to \mathbb{R}$, and a set of actions A^1 such that:*

    (a) For all $a \in A^1$, for all z such that $a \in \psi_i(z)$: $i \in g_y(z)$ and $g_{t,i}(z) = t_i^1$.

    (b) $A^1 \cap A(I_i^*) \neq \emptyset$.

    (c) For all $I_i', I_i'' \in \{I_i : I_i^* \in \psi_i(I_i)\}$.

        i. If $I_i' \in \psi_i(I_i'')$, then $\tilde{t}_i^1(I_i') \geq \tilde{t}_i^1(I_i'')$.

        ii. If I_i' is the penultimate information set in $\psi_i(I_i'')$ and $\tilde{t}_i^1(I_i') > \tilde{t}_i^1(I_i'')$, then $A^1 \cap A(I_i'') \neq \emptyset$.

        iii. If $I_i' \in \psi_i(I_i'')$ and $\tilde{t}_i^0(I_i') > \tilde{t}_i^0(I_i'')$, then $|A(I_i') \setminus A^1| = 1$.

    (d) For all z where $I_i^* \in \psi_i(z)$:

        i. Either: $i \in g_y(z)$ and $g_{t,i}(z) = t_i^1$.

        ii. Or: $i \notin g_y(z)$ and $g_{t,i}(z) = \tilde{t}_i^0(I_i')$, where I_i' is the last information set in $\psi_i(z)$.

Notice what this definition does not require. The going transfer need not be equal across agents. Whether and how much one agent's going transfer changes could depend on other agents' actions. Some agents could face a procedure consistent with the 'either' clause, and other agents could face a procedure consistent with the 'or' clause. Indeed, which procedure an agent faces could depend on other agents' actions.

Theorem 3. *If $(G, (S^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t), then $\tilde{G} \equiv \mathcal{P}(G, (S^\theta)_{\theta \in \Theta})$ is a monotone price mechanism, and $(\tilde{G}, (\tilde{S}^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t), where $(\tilde{S}^\theta)_{\theta \in \Theta}$ is $(S^\theta)_{\theta \in \Theta}$ restricted to $\tilde{G}$.*

The next theorem characterizes the set of OSP-implementable allocation rules. It invokes two additional assumptions.

First, we assume that f_y *admits a finite partition*, which means that we can partition the type space into a finite set of $|N|$-dimensional intervals, with the allocation rule constant within each interval. This assumption is largely technical. It is required because OSP is defined for discrete-time extensive game forms. OSP is not defined for continuous-time auctions, although we can approximate some of them arbitrarily finely.[18]

---

[18]Simon and Stinchcombe (1989) show that discrete time with a very fine grid can be a good proxy for continuous time. However, in their theory, players have perfect information about past activity in the system. Adapting this to our theory, where $\mathcal{G}$ includes all discrete-time game forms with imperfect information, is far from straightforward.

Second, we assume that the lowest type of each agent is never in the allocation, and has a zero transfer. This assumption is borrowed from the Revenue Equivalence Theorem (Myerson, 1981). However, it is a substantive restriction, and rules out certain cases of interest.

Definition 16. f_y **admits a finite partition** *if there exists $K \in \mathbb{N}$ such that, for each i, there exists $\{\theta_i^k\}_{k=1}^K$ such that:*

1. *$\underline{\theta}_i = \theta_i^1 < \theta_i^2 < \ldots < \theta_i^K = \overline{\theta}_i$.*

2. *For all θ_i, θ_i', for all θ_{-i}, if there does not exist k such that $\theta_i \leq \theta_i^k < \theta_i'$, then $f_y(\theta_i, \theta_{-i}) = f_y(\theta_i', \theta_{-i})$*

The use of a single K for all agents is without loss of generality.

All vector inequalities in the following theorem are in the product order. That is, $v \geq v'$ iff for every index i, $v_i \geq v_i'$. Similarly, $v > v'$ iff for every index i, $v_i > v_i'$.

Theorem 4. *Assume that:*

1. *f_y admits a finite partition.*

2. *For all i, for all θ_{-i}, $i \notin f_y(\underline{\theta}_i, \theta_{-i})$.*

   There exists G and f_t such that:

1. *G OSP-implements (f_y, f_t)*

2. *For all i, for all θ_{-i}, $f_{t,i}(\underline{\theta}_i, \theta_{-i}) = 0$*

   if and only if

1. *f_y is monotone.*

2. *For all $A \subseteq N$, for all $\theta_{N \setminus A}$, for*

$$\tilde{\Theta}_A(\theta_{N \setminus A}) \equiv \cap_{i \in A} \text{closure}(\{\theta_A : \forall \theta_{A \setminus i}' \geq \theta_{A \setminus i} : i \notin f_y(\theta_i, \theta_{A \setminus i}', \theta_{N \setminus A})\}) \tag{9}$$

   (a) $\tilde{\Theta}_A(\theta_{N \setminus A})$ is connected.
   (b) There exists $i \in A$ such that, if $\theta_A > \sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$, then $i \in f_y(\theta_A, \theta_{N \setminus A})$.

The sets defined by Equation 9 are join-semilattices.[19] Since their supremum is also the supremum of a finite set of partition coordinates, it is well defined.

---

[19]For a proof, see Lemma 8 in Appendix C.

5.2 Online Advertising Auctions

We now study an online advertising environment, which generalizes Edelman et al. (2007).

There are n bidders, and $n - 1$ advertising positions.[20] Each position has an associated click-through rate α_k, where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_{n-1} > 0$. For convenience, we define position n with $\alpha_n = 0$.

Each bidder's type is a vector, $\theta_i \equiv (\theta_i^k)_{k=1}^n$. A bidder with type θ_i who receives position k and transfer t has utility:

$$u_i(k, t, \theta_i) = \alpha_k \theta_i^k + t \tag{10}$$

The marginal utility of moving to position k from position k', for type θ_i, is

$$m(k, k', \theta_i) \equiv \alpha_k \theta_i^k - \alpha_{k'} \theta_i^{k'} \tag{11}$$

We make the following assumptions on the type space Θ:

A1. *Finite*:

$$|\Theta| < \infty \tag{12}$$

A2. *Higher slots are better*:

$$\forall k \leq n - 1 : \forall \theta \in \Theta : \forall i \in N : m(k, k+1, \theta_i) \geq 0 \tag{13}$$

A3. *Single-crossing*:[21]

$$\forall k \leq n - 2 : \forall \theta, \theta' \in \Theta : \forall i, j \in N : \text{If } m(k, k+1, \theta_i) > m(k, k+1, \theta_j'),$$
$$\text{then } m(k+1, k+2, \theta_i) > m(k+1, k+2, \theta_j'). \tag{14}$$

A1 is a technical assumption to accommodate extensive game forms that move in discrete steps. A2 and A3 are substantive assumptions. Edelman et al. (2007) assume that for all k, k', $\theta_i^k = \theta_i^{k'}$, which entails A2. If $\alpha_1 > \alpha_2 > \ldots > \alpha_{n-1} > 0$, then their assumption also entails A3.

In this environment, the Vickrey-Clarke-Groves (VCG) mechanism selects the efficient allocation. Suppose we number each buyer according to the slot he wins. Then bidder i has VCG payment:

---

[20] It is trivial to extend what follows to fewer than $n - 1$ advertising positions, but doing so would add notation.

[21] This assumption is not identical to the single-crossing assumption in Yenmez (2014). For instance, their condition permits the second inequality in Equation 14 to be weak.

$$-t_i = \sum_{k=i}^{n-1} m(k, k+1, \theta_{k+1}) \tag{15}$$

Edelman et al. (2007) produce a generalized English auction that *ex post* implements the efficient allocation rule in online advertising auctions. The generalized English auction has a unique perfect Bayesian equilibrium in continuous strategies. It is not SP, and therefore is not OSP.

Here we produce an alternative ascending auction that OSP-implements the efficient allocation rule.

Proposition 3. *Assume A1, A2, A3. There exists G that OSP-implements the efficient allocation rule and the VCG payments.*

Proof. We construct G. Set $p_{n-1} := 0$, $A_{n-1} = N$.
 For $l = 1, \ldots, n-1$:

1. Start the price at p_{n-l}.

2. Raise the price in small increments. If the current price is p'_{n-l}, the next price is:

$$p''_{n-l} :=$$
$$\inf_{\theta \in \Theta, i \in N} \{ m(n-l, n-l+1, \theta_i) : m(n-l, n-l+1, \theta_i) > p'_{n-l} \} \tag{16}$$

3. At each price, query each agent in A_{n-l} (in an arbitrary order), giving her the option to quit.

4. At any price p'_{n-l}, if agent i quits, allocate her slot $n-l+1$, and charge every agent in $A_{n-l} \setminus i$ the price p'_{n-l}.

5. Set

$$p_{n-l-1} :=$$
$$\inf_{\theta \in \Theta, i \in N} \{ m(n-l-1, n-l, \theta_i) : m(n-l, n-l+1, \theta_i) \geq p'_{n-l} \} \tag{17}$$

$$A_{n-l-1} := A_{n-l} \setminus i \tag{18}$$

It is an obviously dominant strategy for agent i to quit iff the price in round l is weakly greater than $m(n-l, n-l+1, \theta_i)$.

Consider any round l. Payments from previous rounds are sunk costs. Quitting yields slot $n - l + 1$ at no additional cost, and removes the agent from future rounds.

Consider deviations where the earliest point of departure involves quitting. The current price p'_{n-l} is weakly less than $m(n - l, n - l + 1, \theta_i)$. If the truth-telling strategy has the result that i quits in round l, this outcome is at least as good for i as quitting now. If the truth-telling strategy has the result that i does not quit in round l, then i is charged some amount less than his marginal value for moving up a slot, and the next starting price is $p_{n-l-1} \leq m(n - l - 1, n - l, \theta_i)$, so the argument repeats.

Consider deviations where the earliest point of departure involves staying in. The current price p'_{n-l} is weakly greater than $m(n - l, n - l + 1, \theta_i)$, so this either has the same result as quitting now, or raises i's position at marginal cost weakly above i's marginal utility. This is trivially true for the current round. Consider the next round, $l + 1$. If the starting price p_{n-l-1} is strictly less $m(n - l - 1, n - l, \theta_i)$, then there exists some θ' and j such that $m(n - l - 1, n - l, \theta_i) > m(n - l - 1, n - l, \theta'_j)$. And $m(n - l, n - l + 1, \theta_i) \leq p'_{n-l} \leq m(n - l, n - l + 1, \theta'_j)$, which contradicts A3. Repeating the argument suffices to prove the claim for all rounds $l' \geq l$.

By inspection, this mechanism and the specified strategy profile result in the efficient allocation and the VCG payments.

$\square$

Internet transactions conducted by a central auctioneer raise commitment problems, and bidders may be legitimately concerned about shill bidding. If we consider such auctions as repeated games, reputation can ameliorate commitment problems, but the set of equilibria can be very large and prevent tractable analysis.

Proposition 3 shows that, even if we do not consider such auctions as repeated games, there sometimes exist robust mechanisms that rely only on bilateral commitments. In the case of advertising auctions, the speed of transactions may require bidders to implement their strategies using automata.

5.3 Top Trading Cycles

We now produce an impossibility result for OSP-implementation in a classic matching environment. (Shapley and Scarf, 1974)

There are n agents in the market, each endowed with an indivisible good. An agent's type is a vector $\theta_i \in \mathbb{R}^n$. Θ is the set of all n by n matrices of

real numbers. An outcome assigns one object to each agent. If agent i is assigned object k, he has utility θ_i^k. There are no money transfers.

Following Roth (1982), we assume that the algorithm in question has an arbitrary, fixed way of resolving ties.

Given preferences θ and agents $R \subseteq N$, a *top trading cycle* is a set $\emptyset \subset R' \subseteq R$ whose members can be indexed in a cyclic order:

$$R' = \{i_1, i_2, \ldots, i_r = i_0\} \tag{19}$$

such that each agent i_k likes i_{k+1}'s good more than any other good in R, resolving ties according to the fixed order.

Definition 17. *f is a top trading cycle rule if, for all θ, $f(\theta)$ is equal to the output of the following algorithm:*

1. *Set $R^1 := N$*

2. *For $l = 1, 2, , \ldots$:*

    (a) *Choose some top trading cycle $R' \subseteq R^l$.*
    (b) *Carry out the indicated trades.*
    (c) *Set $R^{l+1} := R^l \setminus R'$.*
    (d) *Terminate if $R^{l+1} = \emptyset$.*

Proposition 4. *If f is a top trading cycle rule, then there exists G that SP-implements f.*

This result is proved in Roth (1982).

Proposition 5. *If $n \geq 3$ and f is a top trading cycle rule, then there does not exist G that OSP-implements f.*

Proof. Saks and Yu (2005) note that SP-implementability is a hereditary property of functions. That is, if f is SP-implementable given domain Θ, then the subfunction $f' = f$ with domain $\Theta' \subseteq \Theta$ is SP-implementable. By inspection, the same is true for OSP-implementability. Thus, to prove Proposition 5, it suffices to produce a subfunction that is not OSP-implementable.

Consider the following subset of $\Theta' \subset \Theta$. Take agents a, b, c, with endowed goods A, B, C. a has only two possible types, θ_a and θ_a', such that

$$\begin{aligned} \text{Either } & B \succ_a C \succ_a A \succ_a \ldots \\ \text{or } & C \succ_a B \succ_a A \succ_a \ldots \end{aligned} \tag{20}$$

25

We make the symmetric assumption for b and c.

We now argue by contradiction. Take any G pruned with respect to the truthful strategy profiles, such that G OSP-implements $f' = f$ for domain Θ'.[22] Consider some history h at which $P(h) = a$ with a non-singleton action set. This cannot come before all such histories for b and c.

Suppose not, and suppose $B \succ_a C$. If a chooses the action corresponding to $B \succ_a C$, and faces opponent strategies corresponding to $C \succ_b A$ and $B \succ_c A$, then a receives good A. If a chooses the action corresponding to $C \succ_a B$, and faces opponent strategies corresponding $A \succ_c B$, then a receives good C. Thus, it is not an obviously dominant strategy to choose the action corresponding to $B \succ_a C$. So a cannot be the first to have a non-singleton action set.

By symmetry, this argument applies to b and c as well. So all of the action sets for a, b, and c are singletons, and G does not OSP-implement f', a contradiction. □

Proposition 5 implies that no implementation of top trading cycles can be deduced to be strategy-proof without contingent reasoning. This is well, since few would claim that the main result of Roth (1982) is obvious.

6   Conclusion

In this paper, we produced a compact definition of obviously strategy-proof mechanisms. This uses a novel notion: We define the earliest points of departure of two strategies. Appendix B shows that this notion is latent in our usual definition of strategy-proofness.

Using a formal model of cognitive limitations, we proved that a strategy is obviously dominant if and only if it can be deduced to be weakly dominant without contingent reasoning. We proved that a choice rule is OSP-implementable if and only if it can be supported by bilateral commitments.

For binary allocation problems, we characterized the OSP mechanisms and the OSP-implementable allocation rules. We produced one possibility result for a case with multi-minded bidders, and one impossibility result for a classic matching algorithm.

Much remains to be done. There are many classic results for SP-implementation, where OSP-implementation is an open question. In this paper, we considered a cognitive model that rules out contingent reasoning. It remains to be

---

[22]Lemma 2 establishes that, if f' is OSP-implementable, then such a pruned G exists.

seen whether there exists a tractable model of agents who can use contingent reasoning, but of limited complexity. Finally, this paper has characterized one natural partial commitment metagame, but there exist others; and it would be valuable to find a way to think rigorously and generally about partial commitment.

References

Ausubel, L. M. (2004). An efficient ascending-bid auction for multiple objects. *American Economic Review*, pages 1452–1475.

Bartal, Y., Gonen, R., and Nisan, N. (2003). Incentive compatible multi unit combinatorial auctions. In *Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge*, pages 72–87. ACM.

Charness, G. and Levin, D. (2009). The origin of the winner's curse: a laboratory study. *American Economic Journal: Microeconomics*, pages 207–236.

Cramton, P. (1998). Ascending auctions. *European Economic Review*, 42(3):745–756.

Crawford, V. P. and Iriberri, N. (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770.

Edelman, B., Ostrovsky, M., and Schwarz, M. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review*, pages 242–259.

Esponda, I. (2008). Behavioral equilibrium in economies with adverse selection. *The American Economic Review*, 98(4):1269–1291.

Esponda, I. and Vespa, E. (2014). Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics*, 6(4):180–202.

Eyster, E. and Rabin, M. (2005). Cursed equilibrium. *Econometrica*, 73(5):1623–1672.

Friedman, E. J. (2002). Strategic properties of heterogeneous serial cost sharing. *Mathematical Social Sciences*, 44(2):145–154.

Friedman, E. J. (2004). Asynchronous learning in decentralized environments: A game-theoretic approach. In *Collectives and the Design of Complex Systems*, pages 133–143. Springer.

Green, J. and Laffont, J.-J. (1977). Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica: Journal of the Econometric Society*, pages 427–438.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, pages 74–91.

Kagel, J. H., Harstad, R. M., and Levin, D. (1987). Information impact and allocation rules in auctions with affiliated private values: A laboratory study. *Econometrica*, 55(6):pp. 1275–1304.

Kagel, J. H. and Levin, D. (1993). Independent private value auctions: Bidder behaviour in first-, second-and third-price auctions with varying numbers of bidders. *The Economic Journal*, pages 868–879.

Milgrom, P. and Segal, I. (2014). Deferred-acceptance auctions and radio spectrum reallocation. *working paper*.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, pages 175–208.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1):58–73.

Piccione, M. and Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20(1):3–24.

Roth, A. E. (1982). Incentive compatibility in a market with indivisible goods. *Economics letters*, 9(2):127–132.

Rothkopf, M. H., Teisberg, T. J., and Kahn, E. P. (1990). Why are Vickrey auctions rare? *Journal of Political Economy*, pages 94–109.

Saks, M. and Yu, L. (2005). Weak monotonicity suffices for truthfulness on convex domains. In *Proceedings of the 6th ACM conference on Electronic commerce*, pages 286–293. ACM.

Savage, L. J. (1954). *The foundations of statistics*. Wiley & Sons, New York.

Shafir, E. and Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive psychology*, 24(4):449–474.

Shapley, L. and Scarf, H. (1974). On cores and indivisibility. *Journal of mathematical economics*, 1(1):23–37.

Simon, L. K. and Stinchcombe, M. B. (1989). Extensive form games in continuous time: Pure strategies. *Econometrica: Journal of the Econometric Society*, pages 1171–1214.

Spence, M. (1974). Competitive and optimal responses to signals: An analysis of efficiency and distribution. *Journal of Economic theory*, 7(3):296–332.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37.

Yenmez, M. B. (2014). Pricing in position auctions and online advertising. *Economic Theory*, 55(1):243–256.

A δ_c has full support

Defining "full support" is a subtle issue. The support of a probability distribution is the smallest closed set whose complement has probability zero. However, our definition of extensive game forms permits $A(h)$ to be uncountably infinite, and does not specify a topology on $A(h)$. The topology we need is defined using utilities:

Definition 18. δ_c *has full support if*

$$\forall i, h, S_i, S_{-i}, d_c, \theta_i : \forall \epsilon > 0 :$$
$$P_{d'_c \sim \delta_c}(|u_i^G(h, S_i, S_{-i}, d'_c, \theta_i) - u_i^G(h, S_i, S_{-i}, d_c, \theta_i)| < \epsilon) > 0 \qquad (21)$$

Consider the values of $u_i^G(h, S_i, S_{-i}, d_c, \theta_i)$ that can be achieved for some realization of d_c. Full support requires that there does not exist an open ball in the achievable utilities that has probability zero under δ_c. Equivalently, the smallest closed set whose complement has probability zero is the entire space of achievable utilities.

B Weak dominance restated using earliest points of departure

Definition 19. *In G for agent i with preferences θ_i, S_i is **weakly dominant including chance realizations** if:*

$$\forall S_i' : \forall S_{-i}' : \forall d_c :$$
$$u_i^G(h_\emptyset, S_i', S_{-i}', d_c, \theta_i) \le u_i^G(h_\emptyset, S_i, S_{-i}', d_c, \theta_i) \tag{22}$$

Proposition 6. *S_i is **weakly dominant including chance realizations** if and only if*

$$\forall S_i' : \forall I_i \in \alpha(S_i, S_i') : \forall h \in I_i : \forall S_{-i}' : \forall d_c :$$
$$u_i^G(h, S_i', S_{-i}', d_c, \theta_i) \le u_i^G(h, S_i, S_{-i}', d_c, \theta_i) \tag{23}$$

Proof.

Lemma 1. *The following statement holds for any S_i. If there exists $h \in I_i, S_{-i}, d_c$ such that h is a subhistory of $z^G(h_\emptyset, S_i, S_{-i}, d_c)$, then for all $h' \in I_i$, there exists S_{-i}', d_c' such that h' is a subhistory of $z^G(h_\emptyset, S_i, S_{-i}', d_c')$.*

This holds because, if there is some history in $h' \in I_i$ that is never reached if S_i is played, then perfect recall requires that h and h' be in different information sets.

First we show that if Equation 22 does not hold, then Equation 23 does not hold. Suppose Equation 22 does not hold. That is

$$\exists S_i', S_{-i}', d_c : u_i^G(\theta_i, h_\emptyset, S_i', S_{-i}', d_c) > u_i^G(\theta_i, h_\emptyset, S_i, S_{-i}', d_c) \tag{24}$$

Thus, $z^G(h_\emptyset, S_i, S_{-i}', d_c) \ne z^G(h_\emptyset, S_i', S_{-i}', d_c)$. So there exists a subhistory h of $z^G(h_\emptyset, S_i, S_{-i}', d_c)$ where $h \in I_i$ and $I_i \in \alpha_i(S_i, S_i')$. $(S_i', I_i, h, S_{-i}', d_c)$ is a counterexample to Equation 23.

Now the other direction. Suppose Equation 23 is false. Then

$$\exists S_i', I_i \in \alpha(S_i, S_i'), h \in I_i, \check{S}_{-i}, \check{d}_c :$$
$$u_i^G(\theta_i, h, S_i', \check{S}_{-i}, \check{d}_c) > u_i^G(\theta_i, h, S_i, \check{S}_{-i}, \check{d}_c) \tag{25}$$

By Lemma 1, we can find some $\hat{S}_i, \hat{d}_c$ that brings us to h. That is, there exists $\hat{S}_i, \hat{d}_c$ such that h is a subhistory of $z^G(h_\emptyset, S_i, \hat{S}_i, \hat{d}_c)$.

Now we specify a synthesis of $(\check{S}_{-i}, \check{d}_c)$ and $(\hat{S}_{-i}, \hat{d}_c)$. Define $(\tilde{S}_{-i}, \tilde{d}_c)$ as follows:

1. If there is a subhistory h' of h such that $h' \in I'_j$ for $j \neq i$, then $\tilde{S}_j(I'_j) = \hat{S}_i(I'_j)$

2. For all other I'_j such that $j \neq i$: $\tilde{S}_j(h') = \check{S}_j(h')$

3. If there is a subhistory h' of h such that $P(h') = c$, then $\tilde{d}_c(h') = \hat{d}_c(h')$

4. For all other h' such that $P(h') = c$: $\tilde{d}_c(h') = \check{d}_c(h')$

Notice that no I'_j can contain both proper subhistories of h and proper superhistories of h, by the perfect recall assumption. j remembers information set I'_j, and can tell whether he has visited it before.

Since $I_i \in \alpha(S_i, S'_i)$, it follows by construction that $z^G(h_\emptyset, S'_i, \tilde{S}_{-i}, \tilde{d}_c) = z^G(h, S'_i, \check{S}_{-i}, \check{d}_c)$ and $z^G(h_\emptyset, S_i, \tilde{S}_{-i}, \tilde{d}_c) = z^G(h, S_i, \check{S}_{-i}, \check{d}_c)$. Thus,

$$
\begin{aligned}
u^G(\theta_i, h_\emptyset, S'_i, \tilde{S}_{-i}, \tilde{d}_c) &= u^G(\theta_i, h, S'_i, \check{S}_{-i}, \check{d}_c) \\
&> u_i^G(\theta_i, h, S_i, \check{S}_{-i}, \check{d}_c) = u_i^G(\theta_i, h_\emptyset, S_i, \tilde{S}_{-i}, \tilde{d}_c)
\end{aligned}
\tag{26}
$$

$\square$

C   Proofs omitted from the main text

C.1   Theorem 1

Proof. First we prove the "if" direction. Fix agent 1 and preferences θ_1. Suppose that S_1 is not obviously dominant in game G. We need to demonstrate that there exists G'' that generates the same experiences as G, such that $\lambda(S_1)$ is not weakly dominant in G''. We proceed by construction.

Let $(S'_1, I_1, h^{\text{inf}}, S^{\text{inf}}_{-1}, d^{\text{inf}}_c, h^{\text{sup}}, S^{\text{sup}}_{-1}, d^{\text{sup}}_c)$ be some tuple that falsifies Equation 4. Using the same technique as in the proof of Proposition 6 (see Appendix B), we can choose $(S_1, S^{\text{inf}}_{-1}, d^{\text{inf}}_c, S'_1, S^{\text{sup}}_{-1}, d^{\text{sup}}_c)$ such that $(S_1, S^{\text{inf}}_{-1}, d^{\text{inf}}_c)$ causes h^{inf} to be on the path of play, and $(S'_1, S^{\text{sup}}_{-1}, d^{\text{sup}}_c)$ causes h^{sup} to be on the path of play.

First, let G' be the same as G, except that all players other than 1 have perfect information: All their information sets are singleton. Since Definition 1 makes no reference to other players' information partitions, G' generates the same experiences for 1.

Second, let G'' be the same as G', except that Player 2 has a new information set at the start of the game.[23] If 2 plays L at this information set, then the game proceeds exactly as in G'. If 2 plays R at this information set, then play proceeds as follows: Let $R = h_1$. Initialize $k = 1$, and apply the following algorithm:

1. $P(h_k) = 1$.

2. $h_k \in I_1^k$, where I_1^k is the (bijected partner of the) kth information set along $\psi_1(I_1)$.[24]

3. If the algorithm has reached I_1 (i.e. if we have reached the point where S_1 and S_1' disagree), then stop the algorithm. Else:

4. Let h_{k+1} be the history where 1, at h_k, played the same action as that specified in $\psi_1(I_1)$.

5. If 1 plays a different action from that specified in $\psi_1(I_1)$, then play proceeds as from any $h' \neq h_k$, where $h' \in I_1^k$.

6. Repeat the above steps for $k + 1$.

When the algorithm stops, call that information set I_1''. This is the bijected partner of I_1 in G, the first point of difference.

Now we define what happens at the history following R which is in I_1''. Call this history h_α. At h_α, if 1 plays $S_1(I_1)$, then play proceeds as though from $h^{\inf}$ in G'. If 1 plays $S_1'(I_1)$, then play proceeds as though from h^{sup} in G'. If 1 plays $a \notin S_1(I_1) \cup S_1'(I_1)$, then 1 proceeds as though from any $h \in I_1$ in G'.

In words, this is what we have just done: We have constructed a G'' where Player 2 has the option to secretly move 1 into a 'parallel' game. In this parallel game, if 1 follows S_1, then 1 is mechanically directed to (a history that is equivalent to) $h^{\inf}$. If 1 follows S_1', then 1 is mechanically directed to h^{sup}. If 1 follows neither S_1 nor S_1', then 1 is put into a part of the game tree that can only lead to experiences for 1 that are also experiences for 1 in game G'. By construction, G'' generates the same experiences for 1 as G'.

___

[23] If $|N| = 1$, so that there is no second player, then all the steps in this proof work using chance moves instead. We use Player 2 here, merely to illustrate that these equivalences still hold if the environment $\mathcal{G}$ is instead defined to be extensive game forms *without* chance moves.

[24] That is, 1's information sets are such that, if 1 has not yet deviated from S_1, then 1 cannot tell if 2 played L or R.

Now we show that $\lambda(S_1)$ is not weakly dominant in G''. To do this, we define (S_{-1}, d_c) as follows: 2 plays R at the start of the game. At histories following $(h_\alpha, S_1(I_1))$, play proceeds according to $(S_{-1}^{\inf}, d_c^{\inf})$. At histories following $(h_\alpha, S_1'(I_1))$, play proceeds according to $(S_{-1}^{\sup}, d_c^{\sup})$. At other histories, it does not matter for our purposes how play proceeds.

Thus, if 1 plays $\lambda(S_1)$, he reaches outcome $g(Z^G(h, S_1, S_{-1}^{\inf}, d_c^{\inf}))$, whereas if he plays $\lambda(S_1')$, he reaches outcome $g(Z^G(h, S_1, S_{-1}^{\sup}, d_c^{\sup}))$, which he strictly prefers. It follows that $\lambda(S_1)$ is not weakly dominant in G'', since

$$u_1^{G''}(h_\emptyset, \lambda(S_1), S_{-1}, d_c, \theta_1) < u_1^{G''}(h_\emptyset, \lambda(S_1'), S_{-1}, d_c, \theta_1) \qquad (27)$$

But notice that we have not yet specified the chance function δ_c for G''. Now we specify that, whenever $P(h') = c$, $\delta_c(h')$ places probability $1 - \epsilon$ on $d_c(h')$, and probability ϵ divided evenly on the other actions at h'.[25] By continuity of $\mathbb{E}[u_1^{G''}]$ in ϵ, there exists $\epsilon > 0$ such that

$$\mathbb{E}_{\delta_c}[u_1^{G''}(h_\emptyset, \lambda(S_1), S_{-1}, d_c, \theta_1)] < \mathbb{E}_{\delta_c}[u_1^{G''}(h_\emptyset, \lambda(S_1'), S_{-1}, d_c, \theta_1)] \qquad (28)$$

which demonstrates that $\lambda(S_1)$ is not weakly dominant in G''.

Now the "only if" direction. Given (G, S_1), suppose some $(G', \lambda(S_1))$ in its equivalence class such that $\lambda(S_1)$ is not weakly dominant in G'. We want to prove that S_1 is not obviously dominant in G.

For some $(S_1', S_{-1}', \delta_c)$:

$$\mathbb{E}_{\delta_c}[u_1^{G'}(h_\emptyset, \lambda(S_1), S_{-1}', d_c, \theta_1)] < \mathbb{E}_{\delta_c}[u_1^{G'}(h_\emptyset, S_1', S_{-1}', d_c, \theta_1)] \qquad (29)$$

It follows that there exists d_c' such that:

$$u_1^{G'}(h_\emptyset, \lambda(S_1), S_{-1}', d_c', \theta_1) < u_1^{G'}(h_\emptyset, S_1', S_{-1}', d_c', \theta_1) \qquad (30)$$

Consider the histories $z^{G'}(h_\emptyset, \lambda(S_1), S_{-1}', d_c')$ and $z^{G'}(h_\emptyset, S_1', S_{-1}', d_c')$. The first difference between them is just after some $h' \in I_1'$, when Player 1 chooses a different action. Let $I_i \equiv \lambda^{-1}(I_i')$. Notice that $I_i \in \alpha(S_1, \lambda^{-1}(S_1'))$.

Since G generates the same experiences as G', there exists some $(\check{S}_{-1}, \check{d}_c)$ such that $g(z^G(h_\emptyset, S_1, \check{S}_{-1}, \check{d}_c)) = g'(z^{G'}(h_\emptyset, \lambda(S_1), S_{-1}', d_c'))$. Similarly, there exists some $(\hat{S}_{-1}, \hat{d}_c)$ such that $g(z^G(h_\emptyset, \lambda^{-1}(S_1'), \hat{S}_{-1}, \hat{d}_c)) = g'(z^{G'}(h_\emptyset, S_1', S_{-1}', d_c'))$.

Both $\psi_1(z^{G'}(h_\emptyset, \lambda(S_1), S_{-1}', d_c'))$ and $\psi_1(z^{G'}(h_\emptyset, S_1', S_{-1}', d_c'))$ include I_1'. Consequently, there exists a subhistory $\check{h}$ of $z^G(h_\emptyset, S_1, \check{S}_{-1}, \check{d}_c)$ and a subhistory $\hat{h}$ of $z^G(h_\emptyset, \lambda^{-1}(S_1'), \hat{S}_{-1}, \hat{d}_c)$ such that $\check{h}, \hat{h} \in I_i$.

---

[25]If there are infinite actions at h', we specify that $\delta_c(h')$ is a linear $(\epsilon, 1-\epsilon)$ combination of any probability measure with full support and an atom of mass on $d_c(h')$.

We have found a deviation $\lambda^{-1}(S_1')$ and an information set $I_i \in \alpha(S_1, \lambda^{-1}(S_1'))$ such that $\check{h}, \hat{h} \in I_i$ and

$$u_1^G(\check{h}, S_1, \check{S}_{-1}, \check{d}_c, \theta_1) < u_1^G(\hat{h}, \lambda^{-1}(S_1'), \hat{S}_{-1}, \hat{d}_c, \theta_1) \tag{31}$$

Thus, S_1 is not obviously dominant in G.

$\square$

C.2   Theorem 2

Proof. First we prove the "only if" direction. Suppose f is OSP-implementable. Let G and $((S_i^{\theta_i})_{i \in N})_{\theta \in \Theta}$ be some mechanism and strategy profiles that OSP-implement f.

Pick any $i \in N$. Each terminal history z in G has some associated experience $\psi_i(z)$ and outcome x. For each information set $I_i \in \mathcal{I}_i$, we associate a unique message $m = \beta(I_i)$. For each action $a \in A(\mathcal{I}_i)$, we associate a unique reply $r = \beta(a)$. We now define

$$\Gamma_i = \{\gamma_i : \neg \exists z \in Z : (\beta(\psi_i(z)), g(z)) = \gamma_i\} \tag{32}$$

This implicitly defines the required bilateral commitment $\hat{\Psi}_0^i$. We now define strategy $\tilde{S}_0^*$, which 'runs' game G. $\tilde{S}_0^*$ is defined as follows:

1. Set $h = h_\emptyset$, the beginning history in G

2. If $P(h) = j$, for $j \in N$, then:

   - Send j message and reply set (m, R) corresponding to the information set such that $h \in I_j$ and the actions $A(I_j)$ in G.

   - Upon receiving reply $r \in R$, set $h := (h, a)$ for a corresponding to r.

   - Go to Step 2.

3. Else if $P(h) = c$, then:

   - Randomize across available actions in $A(h)$ according to measure δ_c in G.

   - Set $h := (h, a)$ for the chosen action.

   - Go to Step 2.

4. If $h \in Z$, then announce outcome $g(z)$ as required by G.

We now show that, for any θ_i, the strategy $\tilde{S}_i^{\theta_i}$, that plays in the bilateral commitment game whatever $S_i^{\theta_i}$ plays in G, is a best response to any $\tilde{S}_{N\setminus i}$ and any $\tilde{S}_0 \in \hat{\Psi}_0^i$.

Suppose not. Then there exists some $\tilde{S}_{N\setminus i}$ and $\tilde{S}_0 \in \hat{\Psi}_0^i$ such that deviation $\tilde{S}_i'$ yields a strictly higher expected payoff than $\tilde{S}_i^{\theta_i}$. We can pick some $\tilde{S}_0$ that never (for any realization) produces $\gamma_i \in \Gamma_i$. Consider the modified strategy $\tilde{S}_0' \in \hat{\Psi}_0^i$, which is equal to $\tilde{S}_0$ with probability $1 - \epsilon$ and is equal to $\tilde{S}_0^*$ with probability ϵ.[26] For small $\epsilon > 0$, $\tilde{S}_i'$ still yields a strictly higher expected payoff than $\tilde{S}_i^{\theta_i}$.

We now produce G' that generates the same experiences for i as G, where $\lambda(S_i^{\theta_i})$ is not weakly dominant. Let G' be the extensive game such that:

1. The extensive game form in G' is the same as the extensive game form in the bilateral commitment game, EXCEPT:

2. Whenever Player 0 is called to play in the bilateral commitment game, chance is called to play in G'.

3. $\tilde{S}_0'$ in the bilateral commitment game is the same as δ_c' in G'.

4. Any subtrees such that all the terminal nodes correspond to $\gamma_i \in \Gamma_i$ are removed from G'.

At every terminal history in G', i's experience $\psi_i'(z')$ and outcome $g'(z')$ correspond to some communication sequence γ_i under $\tilde{S}_0'$, which corresponds to that same communication sequence γ_i under $\tilde{S}_0^*$, which corresponds to some experience $\psi_i(z)$ and that same outcome in G. The same holds in reverse too, since by construction $\tilde{S}_0^*$ can generate every communication sequence not in Γ_i. Moreover, notice that if there was any subhistory h' in G' with an experience that did not have a partner in G, then there would exist a terminal history with that property, and *vice versa*. Thus, G' generates the same experiences for i as G.

Summarizing: G' generates the same experiences for i as G, but $\lambda(S_i^{\theta_i})$ is not weakly dominant in G'. Thus, by Theorem 1, $S_i^{\theta_i}$ is not obviously dominant in G, a contradiction.

Repeating the above procedure for every agent produces the bilateral commitments $(\hat{\Psi}_0^i)_{i \in N}$ that support f.

---

[26]Note that this is *not* a behavioral strategy. We can represent this as the Planner having a personal information set, where he flips an ϵ-weighted coin that decides his state variable at the start of the game.

We now prove the "if" direction. Suppose there exist bilateral commitments $(\hat{\Psi}_0^i)_{i \in N}$ that support f, with requisite strategy profiles $\tilde{S}_0$, $\tilde{S}_N$, and prohibited set Γ_i. Since for all $\tilde{S}_N'$, $\mu(\Gamma_i, \tilde{S}_0, \tilde{S}_N') = 0$, we can pick $\tilde{S}_0$ such that no realization produces $\gamma_i \in \Gamma_i$. We now construct G that OSP-implements f.

As before, let G be the extensive game such that:

1. Whenever Player 0 is called to play in the bilateral commitment game, chance is called to play in G.

2. $\tilde{S}_0$ in the bilateral commitment game is the same as δ_c in G.

3. Any subtrees such that all the terminal nodes correspond to $\gamma_i \in \Gamma_i$ are removed from G.

Once more, let $\beta(\cdot)$ be the bijection from information sets and actions in G to communication sequences $(m_k, R_k)_{k=1}^t$ and responses r under $\tilde{S}_0$.

In G, for all i, θ_i, let $S_i^{\theta_i}$ play the same as $\tilde{S}_i^{\theta_i}$ in the partial commitment game.

Suppose that, for some i, θ_i, $S_i^{\theta_i}$ is not obviously dominant in G. Then, by Theorem 1, we can find G' that generates the same experiences where $\lambda(S_i^{\theta_i})$ is not weakly dominant.

So consider G' and S_{-i}' such that agent i with type θ_i has a strictly profitable deviation. We can equivalently construct strategy $\tilde{S}_0'$, which 'runs' game G'. We omit the details as they are essentially identical to the previous construction. $\tilde{S}_0' \in \hat{\Psi}_0^i$, since every γ_i produced by $\tilde{S}_0'$ corresponds to some $(\psi_i'(z'), g'(z'))$ in G', which corresponds (by λ^{-1}) to some $(\psi_i(z), g(z))$ in G, which corresponds (by β) to some γ_i produced by $\tilde{S}_0$, which (by construction) never produces $\gamma_i \in \Gamma_i$.

Consequently, there exist $\tilde{S}_{N \setminus i}'$ and $\tilde{S}_0' \in \hat{\Psi}_0^i$ such that $\tilde{S}_i^{\theta_i}$ is not a best response, a contradiction. This concludes the proof of the "if" direction. □

C.3 Theorem 3

Proof.

Lemma 2. *Let* $\tilde{G} \equiv \mathcal{P}(G, (S^\theta)_{\theta \in \Theta})$, *and* $(\tilde{S}^\theta)_{\theta \in \Theta}$ *be* $(S^\theta)_{\theta \in \Theta}$ *restricted to* $\tilde{G}$. *If* $(G, (S^\theta)_{\theta \in \Theta})$ *OSP-implements* (f_y, f_t), *then* $(\tilde{G}, (\tilde{S}^\theta)_{\theta \in \Theta})$ *OSP-implements* (f_y, f_t).

Let $\mathbf{S}$ and $\tilde{\mathbf{S}}$ be the sets of possible strategies under G and $\tilde{G}$ respectively. Since $(G, (S^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t), it follows that, for all $S \in (S^\theta)_{\theta \in \Theta}$.

$$\forall i : \forall S_i' \in \mathbf{S}_i : \forall I_i \in \alpha(S_i, S_i') :$$
$$\inf_{h \in I_i, S_{-i}' \in \mathbf{S}_{-i}, d_c} u_i^G(\theta_i, h, S_i, S_{-i}', d_c) \geq \sup_{h \in I_i, S_{-i}' \in \mathbf{S}_{-i}, d_c} u_i^G(\theta_i, h, S_i', S_{-i}', d_c)$$

$$(33)$$

By inspection, $\tilde{\mathbf{S}} \subseteq \mathbf{S}$. Consequently, the above inequality implies

$$\forall i : \forall S_i' \in \tilde{\mathbf{S}}_i : \forall I_i \in \alpha(S_i, S_i') :$$
$$\inf_{h \in I_i, S_{-i}' \in \tilde{\mathbf{S}}_{-i}, d_c} u_i^G(\theta_i, h, S_i, S_{-i}', d_c) \geq \sup_{h \in I_i, S_{-i}' \in \tilde{\mathbf{S}}_{-i}, d_c} u_i^G(\theta_i, h, S_i', S_{-i}', d_c)$$

$$(34)$$

which proves Lemma 2.

Take any $(G, (S^\theta)_{\theta \in \Theta})$ that implements (f_y, f_t). For any history h, we define

$$\Theta_h \equiv \{\theta \in \Theta : h \text{ is a subhistory of } z^G(\emptyset, S^\theta)\} \tag{35}$$

$$\Theta_{h,i} \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_h\} \tag{36}$$

For information set I_i, we define

$$\Theta_{I_i} \equiv \cup_{h \in I_i} \Theta_h \tag{37}$$

$$\Theta_{I_i,i} \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_{I_i}\} \tag{38}$$

$$\Theta_{I_i,i}^1 \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_{I_i} \text{ and } i \in f_y(\theta_i, \theta_{-i})\} \tag{39}$$

$$\Theta_{I_i,i}^0 \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_{I_i} \text{ and } i \notin f_y(\theta_i, \theta_{-i})\} \tag{40}$$

Some observations about this construction:

1. Since player i's strategy depends only on his own type, $\Theta_{I_i,i} = \Theta_{h,i}$ for all $h \in I_i$.

2. $\Theta_{I_i,i} = \Theta^1_{I_i,i} \cup \Theta^0_{I_i,i}$

3. Since SP requires $1_{i \in f_y(\theta)}$ weakly increasing in θ_i, $\Theta^1_{I_i,i}$ dominates $\Theta^0_{I_i,i}$ in the strong set order.

Lemma 3. *Suppose* $(G, (S^\theta)_{\theta \in \Theta})$ *OSP-implements* (f_y, f_t), *where* $G = \langle N, H, P, (\mathcal{I}_i)_{i \in N}, g \rangle$. *For all* i, *for all* $I_i \in \mathcal{I}_i$, *if:*

1. $\theta_i < \theta'_i$

2. $\theta_i \in \Theta^1_{I_i,i}$

3. $\theta'_i \in \Theta^0_{I_i,i}$

then $S^{\theta_i}_i(I_i) = S^{\theta'_i}_i(I_i)$.

Equivalently, for any I_i, *there exists* $a^*_{I_i}$ *such that for all* $\theta_i \in \Theta^1_{I_i,i} \cap \Theta^0_{I_i,i}$, $S^{\theta_i}_i(I_i) = a^*_{I_i}$.

Suppose not. Take $(i, I_i, \theta_i, \theta'_i)$ constituting a counterexample to Lemma 3. Since $\theta_i \in \Theta^1_{I_i,i}$, there exists $h \in I_i$ and S_{-i} such that $i \in g_y(z^G(h, S^{\theta_i}_i, S_{-i}))$. Fix $t_i \equiv g_{t,i}(z^G(h, S^{\theta_i}_i, S_{-i}))$. Since $\theta'_i \in \Theta^0_{I_i,i}$, there exists $h' \in I_i$ and S'_{-i} such that $i \notin g_y(z^G(h', S^{\theta'_i}_i, S'_{-i}))$. Fix $t'_i \equiv g_{t,i}(z^G(h', S^{\theta'_i}_i, S'_{-i}))$. Since $S^{\theta_i}_i(I_i) \neq S^{\theta'_i}_i(I_i)$ and $\theta_i \cup \theta'_i \subseteq \Theta_{I_i,i}$, $I_i \in \alpha(S^{\theta_i}_i, S^{\theta'_i}_i)$. Thus, OSP requires that

$$u_i(\theta_i, h, S^{\theta_i}_i, S_{-i}) \geq u_i(\theta_i, h', S^{\theta'_i}_i, S'_{-i}) \tag{41}$$

which implies

$$\theta_i + t_i \geq t'_i \tag{42}$$

and

$$u_i(\theta'_i, h, S^{\theta_i}_i, S_{-i}) \leq u_i(\theta'_i, h', S^{\theta'_i}_i, S'_{-i}) \tag{43}$$

which implies

$$\theta'_i + t_i \leq t'_i \tag{44}$$

But $\theta'_i > \theta_i$, so

$$\theta'_i + t_i > t'_i \tag{45}$$

a contradiction. This proves Lemma 3. The last statement follows as a corollary of the rest.

Lemma 4. *Suppose $(G, (S^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t) and $\mathcal{P}(G, (S^\theta)_{\theta \in \Theta}) = G$. Take any I_i such that $\Theta^1_{I_i,i} \cap \Theta^0_{I_i,i} \neq \emptyset$, and associated $a^*_{I_i}$.*

1. *If there exists $\theta_i \in \Theta^0_{I_i,i}$ such that $S^{\theta_i}_i(I_i) \neq a^*_{I_i}$, then there exists t^0_i such that:*

   (a) *For all $\theta_i \in \Theta^0_{I_i,i}$ such that $S^{\theta_i}_i(I_i) \neq a^*_{I_i}$, for all $h \in I_i$, for all S_{-i}, $g_{t,i}(z^G(h, S^{\theta_i}_i, S_{-i})) = t^0_i$.*

   (b) *For all $\theta_i \in \Theta_{I_i,i}$ such that $S^{\theta_i}_i(I_i) = a^*_{I_i}$, for all $h \in I_i$, for all S_{-i}, if $i \notin g_y(z^G(h, S^{\theta_i}_i, S_{-i}))$, then $g_{t,i}(z^G(h, S^{\theta_i}_i, S_{-i})) = t^0_i$.*

2. *If there exists $\theta_i \in \Theta^1_{I_i,i}$ such that $S^{\theta_i}_i(I_i) \neq a^*_{I_i}$, then there exists t^1_i such that:*

   (a) *For all $\theta_i \in \Theta^1_{I_i,i}$ such that $S^{\theta_i}_i(I_i) \neq a^*_{I_i}$, for all $h \in I_i$, for all S_{-i}, $g_{t,i}(z^G(h, S^{\theta_i}_i, S_{-i})) = t^1_i$.*

   (b) *For all $\theta_i \in \Theta_{I_i,i}$ such that $S^{\theta_i}_i(I_i) = a^*_{I_i}$, for all $h \in I_i$, for all S_{-i}, if $i \in g_y(z^G(h, S^{\theta_i}_i, S_{-i}))$, then $g_{t,i}(z^G(h, S^{\theta_i}_i, S_{-i})) = t^1_i$.*

Take any type $\theta'_i \in \Theta^0_{I_i,i}$ such that $S^{\theta'_i}_i(I_i) \neq a^*_{I_i}$. Take any type $\theta''_i \in \Theta^0_{I_i,i}$ such that $S^{\theta''_i}_i(I_i) = a^*_{I_i}$. (By $\Theta^1_{I_i,i} \cap \Theta^0_{I_i,i} \neq \emptyset$ there exists at least one such type.) Notice that $I_i \in \alpha(S^{\theta'_i}_i, S^{\theta''_i}_i)$.

By Lemma 3, $\theta'_i \notin \Theta^1_{I_i,i}$, and the game is pruned. Thus,

$$\forall h \in I_i : \forall S_{-i} : i \notin g_y(z^G(h, S^{\theta'_i}_i, S_{-i})) \tag{46}$$

Since $\theta''_i \in \Theta^0_{I_i,i}$,

$$\exists h \in I_i : \exists S_{-i} : i \notin g_y(z^G(h, S^{\theta''_i}_i, S_{-i})) \tag{47}$$

OSP requires that type θ'_i does not want to (inf-sup) deviate. Thus,

$$\inf_{h \in I_i, S_{-i}} g_{t,i}(z^G(h, S^{\theta'_i}_i, S_{-i})) \geq$$
$$\sup_{h \in I_i, S_{-i}} \{g_{t,i}(z^G(h, S^{\theta''_i}_i, S_{-i})) : i \notin g_y(z^G(h, S^{\theta''_i}_i, S_{-i}))\} \tag{48}$$

OSP also requires that type θ''_i does not want to (inf-sup) deviate. This implies

$$\inf_{h \in I_i, S_{-i}} \{g_{t,i}(z^G(h, S_i^{\theta_i''}, S_{-i})) : i \notin g_y(z^G(h, S_i^{\theta_i''}, S_{-i}))\}$$

$$\geq \sup_{h \in I_i, S_{-i}} g_{t,i}(z^G(h, S_i^{\theta_i'}, S_{-i})) \tag{49}$$

The RHS of Equation 48 is weakly greater than the LHS of Equation 49. The RHS of Equation 49 is weakly greater than the LHS of Equation 48. Consequently all four terms are equal. Moreover, this argument applies to every $\theta_i' \in \Theta_{I_i,i}^0$ such that $S_i^{\theta_i'}(I_i) \neq a_{I_i}^*$, and every $\theta_i'' \in \Theta_{I_i,i}^0$ such that $S_i^{\theta_i''}(I_i) = a_{I_i}^*$. Since the game is pruned, θ_i'' satisfies (1b) iff $\theta_i'' \in \Theta_{I_i,i}^0$ and $S_i^{\theta_i''}(I_i) = a_{I_i}^*$. This proves part 1 of Lemma 4. Part 2 follows by symmetry; we omit the details since they involve only small notational changes to the above argument.

Lemma 5. *Suppose $(G, (S^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t) and $\mathcal{P}(G, (S^\theta)_{\theta \in \Theta}) = G$. Take any I_i such that $\Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0 \neq \emptyset$, and associated $a_{I_i}^*$. Let t_i^1 and t_i^0 be defined as before.*

1. *If there exists $\theta_i \in \Theta_{I_i,i}^0$ such that $S_i^{\theta_i}(I_i) \neq a_{I_i}^*$, then for all $(h \in I_i, S_i, S_{-i})$, if $i \in g_y(z^G(h, S_i\,S_{-i}))$, then $g_{t,i}(z^G(h, S_i\,S_{-i})) \leq t_i^0 - \sup\{\theta_i \in \Theta_{I_i,i}^0 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\}$.*

2. *If there exists $\theta_i \in \Theta_{I_i,i}^1$ such that $S_i^{\theta_i}(I_i) \neq a_{I_i}^*$, then for all $(h \in I_i, S_i, S_{-i})$, if $i \notin g_y(z^G(h, S_i\,S_{-i}))$, then $g_{t,i}(z^G(h, S_i\,S_{-i})) \leq \inf\{\theta_i \in \Theta_{I_i,i}^1 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\} + t_i^1$.*

Suppose that part 1 of Lemma 5 does not hold. Fix $(h \in I_i, S_i, S_{-i})$ such that $i \in g_y(z^G(h, S_i\,S_{-i}))$ and $g_{t,i}(z^G(h, S_i, S_{-i})) > t_i^0 - \sup\{\theta_i \in \Theta_{I_i,i}^0 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\}$. Since G is pruned, we can find some $\theta_i' \in \Theta_{I_i,i}$ such that for every $\tilde{I}_i \in \{I_i' \in \mathcal{I}_i : I_i \text{ occurs in } \psi(I_i')\}$, $S_i^{\theta_i'}(\tilde{I}_i) = S_i(\tilde{I}_i)$. Fix that θ_i'.

Fix $\theta_i'' \in \Theta_{I_i,i}^0$ such that $S_i^{\theta_i}(I_i) \neq a_{I_i}^*$ and $\theta_i'' \geq \sup\{\theta_i \in \Theta_{I_i,i}^0 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\} - \epsilon$. Since G is pruned and $\theta_i'' \notin \Theta_{I_i,i}^1$ (by Lemma 3), it must be that $S_i^{\theta_i''}(I_i) \neq S_i^{\theta_i'}(I_i)$.

By construction, $I_i \in \alpha(S_i^{\theta_i'}, S_i^{\theta_i''})$.

OSP requires that, for all $h'' \in I_i, S_{-i}''$:

$$u_i(\theta_i'', h'', S_i^{\theta_i''}, S_{-i}'') \geq u_i(\theta_i'', h, S_i^{\theta_i'}, S_{-i}) \tag{50}$$

which entails

$$t_i^0 \geq \theta_i'' + g_{t,i}(z^G(h, S_i, S_{-i})) \tag{51}$$

which entails

$$t_i^0 - \sup\{\theta_i \in \Theta_{I_i,i}^0 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\} + \epsilon \geq g_{t,i}(z^G(h, S_i, S_{-i})) \tag{52}$$

But, by hypothesis,

$$t_i^0 - \sup\{\theta_i \in \Theta_{I_i,i}^0 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\} < g_{t,i}(z^G(h, S_i, S_{-i})) \tag{53}$$

Since this argument holds for all $\epsilon > 0$, we can pick ϵ small enough to create a contradiction. This proves part 1 of Lemma 5. Part 2 follows by symmetry.

Lemma 6. *Suppose $(G, (S^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t) and $\mathcal{P}(G, (S^\theta)_{\theta \in \Theta}) = G$. Take any I_i such that $|\Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0| > 1$ and associated $a_{I_i}^*$.*

1. *If there exists $\theta_i \in \Theta_{I_i,i}^0$ such that $S_i^{\theta_i}(I_i) \neq a_{I_i}^*$, then for all $\theta_i' \in \Theta_{I_i,i}^1$, $S_i^{\theta_i'}(I_i) = a_{I_i}^*$.*

2. *If there exists $\theta_i \in \Theta_{I_i,i}^1$ such that $S_i^{\theta_i}(I_i) \neq a_{I_i}^*$, then for all $\theta_i' \in \Theta_{I_i,i}^0$, $S_i^{\theta_i'}(I_i) = a_{I_i}^*$.*

Suppose part 1 of Lemma 6 does not hold. Fix I_i, and choose $\theta_i' < \theta_i''$ such that $\theta_i' \cup \theta_i'' \subseteq \Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0$. Fix $\theta_i''' \in \Theta_{I_i,i}^1$ such that $S_i^{\theta_i'''}(I_i) \neq a_{I_i}^*$. By Lemma 3, if $\theta_i''' \in \Theta_{I_i,i}^0$, then $S_i^{\theta_i'''}(I_i) = a_{I_i}^*$, a contradiction. Thus, $\theta_i''' \in \Theta_{I_i,i}^1 \setminus \Theta_{I_i,i}^0$, and since $\Theta_{I_i,i}^1$ dominates $\Theta_{I_i,i}^0$ in the strong set order, $\theta_i'' < \theta_i'''$.

Since $\theta_i' \in \Theta_{I_i,i}^1$, there exists $h' \in I_i$ and θ_{-i}' such that $(\theta_i', \theta_{-i}') \in \Theta_{I_i}$ and $i \in g_y(z^G(h', S_i^{\theta_i'}, S_{-i}^{\theta_{-i}'}))$. By Lemma 4, there exists $a_i \in A_i(I_i)$ such that $a_i \neq a_{I_i}^*$ and choosing a_i ensures $i \notin y$ and $t_i = t_i^0$. Thus, by G SP

$$\theta_i' + g_{t,i}(z^G(h', S_i^{\theta_i'}, S_{-i}^{\theta_{-i}'})) \geq t_i^0 \tag{54}$$

By $\theta_i'' \in \Theta_{I_i,i}^0$, there exists $h'' \in I_i$ and θ_{-i}'' such that $i \notin g_y(z^G(h'', S_i^{\theta_i''}, S_{-i}^{\theta_{-i}''}))$. By Lemma 4

$$g_{t,i}(z^G(h'', S_i^{\theta_i''}, S_{-i}^{\theta_{-i}''})) = t_i^0 \tag{55}$$

By G SP, $i \in g_y(z^G(h', S_i^{\theta_i'''}, S_{-i}^{\theta_{-i}'}))$ and $g_{t,i}(z^G(h', S_i^{\theta_i'''}, S_{-i}^{\theta_{-i}'})) = g_{t,i}(z^G(h', S_i^{\theta_i'}, S_{-i}^{\theta_{-i}'}))$.

Notice that $I_i \in \alpha(S_i^{\theta_i''}, S_i^{\theta_i'''})$. Thus, OSP requires that θ_i'' does not want to (inf-sup) deviate to θ_i''''s strategy, which entails:

$$g_{t,i}(z^G(h'', S_i^{\theta_i''}, S_{-i}^{\theta_{-i}''})) \geq \theta_i'' + g_{t,i}(z^G(h', S_i^{\theta_i'''}, S_{-i}^{\theta_{-i}'})) \tag{56}$$

$$\begin{aligned} t_i^0 &\geq \theta_i'' + g_{t,i}(z^G(h', S_i^{\theta_i'''}, S_{-i}^{\theta_{-i}'})) \\ &> \theta_i' + g_{t,i}(z^G(h', S_i^{\theta_i'}, S_{-i}^{\theta_{-i}'})) \end{aligned} \tag{57}$$

which contradicts Equation 54.

Part 2 follows by symmetry. This proves Lemma 6.

Lemma 7. *Suppose $(G, (S^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t) and $\mathcal{P}(G, (S^\theta)_{\theta \in \Theta}) = G$.*

For all I_i, if $|\Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0| \leq 1$ and $|A(I_i)| \geq 2$, then there exists t_i^1 and t_i^0 such that:

1. *For all $\theta_i \in \Theta_{I_i,i}$, $h \in I_i$, S_{-i}:*

   (a) *If $i \notin g_y(z^G(h, S_i^{\theta_i}, S_{-i}))$ then $g_{t,i}(z^G(h, S_i^{\theta_i}, S_{-i})) = t_i^0$*

   (b) *If $i \in g_y(z^G(h, S_i^{\theta_i}, S_{-i}))$ then $g_{t,i}(z^G(h, S_i^{\theta_i}, S_{-i})) = t_i^1$*

2. *If $|\Theta_{I_i,i}^1| > 0$ and $|\Theta_{I_i,i}^0| > 0$, then $t_i^1 = -\inf\{\theta_i \in \Theta_{I_i,i}^1\} + t_i^0$*

By G pruned and $|A(I_i)| \geq 2$, $\Theta_{I_i,i} \neq \emptyset$.

Consider the case where $\Theta_{I_i,i}^1 = \emptyset$. Pick some $\theta_i' \in \Theta_{I_i,i}^0$ and some $h' \in I_i, S_{-i}'$. Fix $t_i^0 \equiv g_{t,i}(z^G(h', S_i^{\theta_i'}, S_{-i}'))$. Suppose there exists some $(\theta_i'', \theta_{-i}'') \in \Theta_{I_i}$ such that $f_{t,i}(\theta_i'', \theta_{-i}'') = t_i^{0''} \neq t_i^0$. Pick $h'' \in I_i$ in the terminal history $z^G(\emptyset, S_i^{\theta_i''}, S_{-i}^{\theta_{-i}''})$. By Equation 8, for all $\theta_i \in \Theta_{I_i,i}$, $f_{t,i}(\theta_i, \theta_{-i}'') = t_i^{0''}$. By G pruned and $|A(I_i)| \geq 2$, we can pick $\theta_i''' \in \Theta_{I_i,i}^0$ such that $S_i^{\theta_i'''}(I_i) \neq S_i^{\theta_i'}(I_i)$. Notice that $I_i \in \alpha(S_i^{\theta_i'''}(I_i), S_i^{\theta_i'})$. If $t_i^{0''} > t_i^0$, then

$$u_i(\theta_i', h', S_i^{\theta_i'}, S_{-i}') = t_i^0 < t_i^{0''} = u_i(\theta_i', h'', S_i^{\theta_i'''}, S_{-i}'') \tag{58}$$

so $(G, S_i^{\theta_i'})$ is not OSP for (i, θ_i'). If $t_i^{0''} < t_i^0$, then

$$u_i(\theta_i''', h', S_i^{\theta_i'}, S_{-i}') = t_i^0 > t_i^{0''} = u_i(\theta_i''', h'', S_i^{\theta_i'''}, S_{-i}'') \tag{59}$$

42

so $(G, S_i^{\theta_i'''})$ is not OSP for (i, θ_i'''). By contradiction, this proves Lemma 7 for this case. A symmetric argument proves Lemma 7 for the case where $\Theta_{I_i,i}^0 = \emptyset$.

Note that, if Lemma 7 holds at some information set I_i, it holds at all information sets I_i' that follow I_i. Thus, we need only consider some earliest information set I_i^* at which $|\Theta_{I_i^*,i}^1 \cap \Theta_{I_i^*,i}^0| \leq 1$ and $|A(I_i^*)| \geq 2$.

Now we consider the case where $\Theta_{I_i^*,i}^1 \neq \emptyset$ and $\Theta_{I_i^*,i}^0 \neq \emptyset$.

At every prior information set I_i prior to I_i^*, $|\Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0| \geq 2$. Since $\Theta_{I_i^*,i}^1 \neq \emptyset$ and $\Theta_{I_i^*,i}^0 \neq \emptyset$, by Lemma 6, I_i^* is reached by some interval of types all taking the same action. Thus $\sup\{\theta_i \in \Theta_{I_i^*,i}^0\} = \inf\{\theta_i \in \Theta_{I_i^*,i}^1\}$.

Fix $\hat{\theta}_i \in \Theta_{I_i^*,i}^0$ such that $\hat{\theta}_i \geq \sup\{\theta_i \in \Theta_{I_i^*,i}^0\} - \epsilon$. Choose corresponding $\hat{h} \in I_i^*$ and $\hat{\theta}_{-i} \in \Theta_{I_i^*,-i}$ such that $i \notin g_y(z^G(\hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}))$. Define $t_i^0 \equiv g_{t,i}(z^G(\hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}))$.

Fix $\check{\theta}_i \in \Theta_{I_i^*,i}^1$ such that $\check{\theta}_i \leq \inf\{\theta_i \in \Theta_{I_i^*,i}^1\} + \epsilon$. Choose corresponding $\check{h} \in I_i^*$ and $\check{\theta}_{-i} \in \Theta_{I_i^*,-i}$ such that $i \in g_y(z^G(\check{h}, S_i^{\check{\theta}_i}, S_{-i}^{\check{\theta}_{-i}}))$. Define $t_i^1 \equiv g_{t,i}(z^G(\check{h}, S_i^{\check{\theta}_i}, S_{-i}^{\check{\theta}_{-i}}))$.

Suppose there exists some $(\theta_i', \theta_{-i}') \in \Theta_{I_i^*}$ such that $i \notin f_y(\theta_i', \theta_{-i}')$ and $f_{t,i}(\theta_i', \theta_{-i}') = t_i^{0'} \neq t_i^0$. Since $\sup\{\theta_i \in \Theta_{I_i^*,i}^0\} = \inf\{\theta_i \in \Theta_{I_i^*,i}^1\}$, it follows that for all $\theta_{-i} \in \Theta_{I_i^*,-i}$, $\inf\{\theta_i : i \in f_y(\theta_i, \theta_{-i})\} = \inf\{\theta_i \in \Theta_{I_i^*,i}^1\}$. Thus, by Equation 8, for all $\theta_i \in \Theta_{I_i^*,i}$: $f_{t,i}(\theta_i, \theta_{-i}') = -1_{i \in f_y(\theta_i, \theta_{-i}')} \inf\{\theta_i \in \Theta_{I_i^*,i}^1\} + t_i^{0'}$. Fix $h' \in I_i^*$, in the terminal history $z^G(\emptyset, S_i^{\theta_i'}, S_{-i}^{\theta_{-i}'})$.

By $A(I_i) \geq 2$, we can pick some $\theta_i'' \in \Theta_{I_i^*,i}$ such that $S_i^{\theta_i''}(I_i) \neq S_i^{\hat{\theta}_i}(I_i)$. Notice that $I_i^* \in \alpha(S_i^{\theta_i''}, S_i^{\hat{\theta}_i})$. Either $\theta_i'' \in \Theta_{I_i^*}^0$ or $\theta_i'' \in \Theta_{I_i^*}^1 \setminus \Theta_{I_i^*}^0$. Suppose $\theta_i'' \in \Theta_{I_i^*}^0$. Suppose $t_i^{0'} > t_i^0$. By OSP,

$$u_i(\hat{\theta}_i, \hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}) \geq u_i(\hat{\theta}_i, h', S_i^{\theta_i''}, S_{-i}^{\theta_{-i}'}) \tag{60}$$

which entails

$$
\begin{aligned}
t_i^0 &\geq t_i^{0'} + 1_{i \in f_y(\theta_i'', \theta_{-i}')}(\hat{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*,i}^1\}) \\
&\geq t_i^{0'} + 1_{f_y(\theta_i'', \theta_{-i}') \in Y_i}(-\epsilon) \\
&\geq t_i^{0'} - \epsilon
\end{aligned} \tag{61}
$$

and we can pick ϵ small enough to constitute a contradiction. Suppose $t_i^{0'} < t_i^0$. By OSP

$$u_i(\theta_i'', \hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}) \leq u_i(\theta_i'', h', S_i^{\theta_i''}, S_{-i}^{\theta_{-i}'}) \tag{62}$$

which entails

$$t_i^0 \leq t_i^{0'} + 1_{i \in f_y(\theta_i'', \theta_{-i}')}(\theta_i'' - \inf\{\theta_i \in \Theta_{I_i^*,i}^1\})$$
$$= t_i^{0'} + 1_{i \in f_y(\theta_i'', \theta_{-i}')}(\theta_i'' - \sup\{\theta_i \in \Theta_{I_i^*,i}^0\}) \leq t_i^{0'} \tag{63}$$

which is a contradiction.

The case that remains is $\theta_i'' \in \Theta_{I_i^*}^1 \setminus \Theta_{I_i^*}^0$. Then $i \in f_y(\theta_i'', \theta_{-i}')$ and $f_{t,i}(\theta_i'', \theta_{-i}') = -\inf\{\theta_i \in \Theta_{I_i^*,i}^1\} + t_i^{0'}$. Suppose $t_i^{0'} > t_i^0$. OSP requires:

$$u_i(\hat{\theta}_i, \hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}) \geq u_i(\hat{\theta}_i, h', S_i^{\theta_i''}, S_{-i}^{\theta_{-i}'}) \tag{64}$$

which entails

$$t_i^0 \geq \hat{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*,i}^1\} + t_i^{0'}$$
$$\geq t_i^{0'} - \epsilon \tag{65}$$

and we can pick ϵ small enough to constitute a contradiction.

Suppose $t_i^{0'} < t_i^0$. Since $S_i^{\theta_i''}(I_i) \neq S_i^{\hat{\theta}_i}(I_i)$, either $S_i^{\check{\theta}_i}(I_i) \neq S_i^{\hat{\theta}_i}(I_i)$ or $S_i^{\check{\theta}_i}(I_i) \neq S_i^{\theta_i''}(I_i)$. Moreover, $f_{t,i}(\check{\theta}_i, \theta_i') = -1_{f_y(\check{\theta}_i, \theta_{-i}') \in Y_i} \inf\{\theta_i \in \Theta_{I_i^*,i}^1\} + t_i^{0'}$.

Suppose $S_i^{\check{\theta}_i}(I_i) \neq S_i^{\hat{\theta}_i}(I_i)$. OSP requires:

$$u_i(\check{\theta}_i, h', S_i^{\check{\theta}_i}, S_{-i}^{\theta_{-i}'}) \geq u_i(\check{\theta}_i, \hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}) \tag{66}$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta_{-i}')}(\check{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*,i}^1\}) + t_i^{0'} \geq t_i^0 \tag{67}$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta_{-i}')}\epsilon + t_i^{0'} \geq t_i^0 \tag{68}$$

and we can pick ϵ small enough to yield a contradiction. Suppose $S_i^{\check{\theta}_i}(I_i) \neq S_i^{\theta_i''}(I_i)$. By Equation 8, $f_{t,i}(\check{\theta}_i, \theta_{-i}') = -1_{i \in f_y(\check{\theta}_i, \theta_{-i}')} \inf\{\theta_i \in \Theta_{I_i^*,i}^1\} + t_i^{0'}$, and $f_{t,i}(\theta_i'', \hat{\theta}_{-i}) = -\inf\{\theta_i \in \Theta_{I_i^*,i}^1\} + t_i^0$. OSP requires

$$u_i(\check{\theta}_i, h', S_i^{\check{\theta}_i}, S_{-i}^{\theta_{-i}'}) \geq u_i(\check{\theta}_i, \hat{h}, S_i^{\theta_i''}, S_{-i}^{\hat{\theta}_{-i}}) \tag{69}$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta'_{-i})}(\check{\theta}_i - \inf\{\theta_i \in \Theta^1_{I^*_i,i}\}) + t^{0'}_i \geq (\check{\theta}_i - \inf\{\theta_i \in \Theta^1_{I^*_i,i}\}) + t^0_i \quad (70)$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta'_{-i})}\epsilon + t^{0'}_i \geq \epsilon + t^0_i \quad (71)$$

which entails

$$t^{0'}_i \geq t^0_i \quad (72)$$

a contradiction. By the above argument, for all I_i satisfying the assumptions of Lemma 7, there is a unique transfer t^0_i for all terminal histories z passing through I_i such that $i \notin g_y(z)$. Equation 8 thus implies that there is a unique transfer t^1_i for all terminal histories z passing through I_i such that $i \in g_y(z)$. Moreover, $t^1_i = -\inf\{\theta_i \in \Theta^1_{I_i,i}\} + t^0_i$. This proves Lemma 7.

Now to bring this all together.[27] Take any $(G, (S^\theta)_{\theta \in \Theta})$ that OSP-implements (f_y, f_t). Define $\tilde{G} \equiv \mathcal{P}(G, (S^\theta)_{\theta \in \Theta})$ and $(\tilde{S}^\theta)_{\theta \in \Theta}$ as $(S^\theta)_{\theta \in \Theta}$ restricted to $\tilde{G}$. By Lemma 2, $(\tilde{G}, (\tilde{S}^\theta)_{\theta \in \Theta})$ OSP-implements (f_y, f_t).

We now characterize $(\tilde{G}, (\tilde{S}^\theta)_{\theta \in \Theta})$. For any player i, consider any information set I^*_i such that $|A(I^*_i)| \geq 2$, and for all prior information sets $I'_i \in \psi_i(I^*_i) \setminus I^*_i$, $|A(I'_i)| = 1$. By Lemma 3, there is a unique action $a^*_{I^*_i}$ taken by all types in $\Theta^1_{I^*_i,i} \cap \Theta^0_{I^*_i,i}$.

Either $|\Theta^1_{I^*_i,i} \cap \Theta^0_{I^*_i,i}| > 1$ or $|\Theta^1_{I^*_i,i} \cap \Theta^0_{I^*_i,i}| \leq 1$.

If $|\Theta^1_{I^*_i,i} \cap \Theta^0_{I^*_i,i}| > 1$, then by Lemma 6, $\tilde{G}$ pruned and $|A(I_i)| \geq 2$,

1. EITHER: There exists $\theta_i \in \Theta^0_{I^*_i,i}$ such that $S^{\theta_i}_i(I^*_i) \neq a^*_{I^*_i}$, and for all $\theta'_i \in \Theta^1_{I^*_i,i}$, $S^{\theta'_i}_i(I^*_i) = a^*_{I^*_i}$.

2. OR: There exists $\theta_i \in \Theta^1_{I^*_i,i}$ such that $S^{\theta_i}_i(I^*_i) \neq a^*_{I^*_i}$, and for all $\theta'_i \in \Theta^0_{I^*_i,i}$, $S^{\theta'_i}_i(I^*_i) = a^*_{I^*_i}$.

In the first case, then by Lemma 4, there is some t^0_i such that, for all (S_i, S_{-i}), for all $h \in I_i$, if $i \notin g_y(z^G(h, S_i, S_{-i}))$, then $g_{t,i}(z^G(h, S_i, S_{-i}) = t^0_i$. Moreover, we can define a 'going price' at all information sets I'_i such that $I^*_i \in \psi_i(I'_i)$:

---

[27]To any reader who has made it this far, thank you!

45

$$\tilde{t}_i^1(I_i') \equiv \min_{I_i'' \in \psi_i(I_i')} [t_i^0 - \sup\{\theta_i \in \Theta_{I_i'',i}^0 : S_i^{\theta_i}(I_i'') \neq a_{I_i''}^*\}] \tag{73}$$

Notice that this function falls monotonically as we move along the game tree; for any I_i', I_i'' such that $I_i' \in \psi_i(I_i'')$, $\tilde{t}_i^1(I_i') \geq \tilde{t}_i^1(I_i'')$. Moreover, by construction, at any I_i', I_i'' such that I_i'' is the immediate successor of I_i' in i's experience, if $\tilde{t}_i^1(I_i') > \tilde{t}_i^1(I_i'')$, then there exists $a \in A(I_i'')$ that yields $y \notin Y_i$, and by Lemma 4 this yields transfer t_i^0.

Lemma 5 and SP together imply that, at any terminal history z if the going price at i's last information set I_i' was some $\tilde{t}_i^1(I_i')$, and $i \in g_y(z)$, then $g_{t,i}(z) = \tilde{t}_i^1(I_i')$. If $g_{t,i}(z) < \tilde{t}_i^1(I_i')$, then θ_i such that $t_i^0 - \tilde{t}_i^1(I_i') < \theta_i < t_i^0 - g_{t,i}(z)$ would have a profitable deviation at the last information set where $\tilde{t}_i^1$ fell.

In the second case, then by Lemma 4, there is some t_i^1 such that, for all (S_i, S_{-i}), for all $h \in I_i^*$, if $i \in g_y(z^G(h, S_i, S_{-i}))$, then $g_{t,i}(z^G(h, S_i, S_{-i})) = t_i^1$. Moreover, we can define a 'going price' at all information sets I_i' such that $I_i^* \in \psi_i(I_i')$:

$$\tilde{t}_i^0(I_i') \equiv \min_{I_i'' \in \psi_i(I_i')} [t_i^1 + \inf\{\theta_i \in \Theta_{I_i'',i}^0 : S_i^{\theta_i}(I_i'') \neq a_{I_i''}^*\}] \tag{74}$$

Once more,

1. This function falls monotonically as we move along the game tree.

2. At any I_i', I_i'' such that I_i'' is the immediate successor of I_i' in i's experience, if $\tilde{t}_i^0(I_i') > \tilde{t}_i^0(I_i'')$, then there exists $a \in A(I_i'')$ that yields $y \in Y_i$, and transfer t_i^1.

3. For any z, if the going price at i's last information set I_i' was some $\tilde{t}_i^0(I_i')$, and $i \notin g_y(z)$, then $g_{t,i}(z) = \tilde{t}_i^0(I_i')$.

Part (1.c.iii) and (2.c.iii) of Theorem 3 follow from Lemma 7. The above constructions suffice to prove Theorem 3 for cases where $|\Theta_{I_i^*,i}^1 \cap \Theta_{I_i^*,i}^0| > 1$. Cases where $|\Theta_{I_i^*,i}^1 \cap \Theta_{I_i^*,i}^0| \leq 1$ are dealt with by Lemma 7. $\square$

C.4 Theorem 4

Proof. Consider the sets used to construct $\tilde{\Theta}_A(\theta_{N \backslash A})$.

$$\{\theta_A : \forall \theta_{A \backslash i}' \geq \theta_{A \backslash i} : i \notin f_y(\theta_i, \theta_{A \backslash i}', \theta_{N \backslash A})\} \tag{75}$$

These are the type profiles $\theta_A = (\theta_i, \theta_{A\backslash i})$ such that, if all agents in $A \backslash i$ have types *at least* as high as $\theta_{A\backslash i}$ and all agents in $N \backslash A$ have types $\theta_{N\backslash A}$, then the allocation rule requires that type θ_i is not satisfied.

Lemma 8. *For all $A \subseteq N$, for all $\theta_{N\backslash A}$, $\tilde{\Theta}_A(\theta_{N\backslash A})$ is a join-semilattice with respect to the product order on Θ_A.*

Take any $\theta''_A, \theta'''_A \in \tilde{\Theta}_A(\theta_{N\backslash A})$. We want to show that $\theta''_A \vee \theta'''_A \in \tilde{\Theta}_A(\theta_{N\backslash A})$. For all $i \in A$,

$$\theta''_A, \theta'''_A \in \text{closure}(\{\theta_A : \forall \theta'_{A\backslash i} \geq \theta_{A\backslash i} : i \notin f_y(\theta_i, \theta'_{A\backslash i}, \theta_{N\backslash A})\}) \qquad (76)$$

The set on the RHS is upward-closed with respect to the product order on $\theta_{A\backslash i}$.

Consider $\check{\theta}_A \equiv \theta''_A \vee \theta'''_A$. Its ith element has the property: $\check{\theta}_i = \max\{\theta''_i, \theta'''_i\}$. WLOG, suppose $\theta''_i \geq \theta'''_i$. Then, since $\check{\theta}_{A\backslash i} \geq \theta''_{A\backslash i}$,

$$\check{\theta}_A \in \text{closure}(\{\theta_A : \forall \theta'_{A\backslash i} \geq \theta_{A\backslash i} : i \notin f_y(\theta_i, \theta'_{A\backslash i}, \theta_{N\backslash A})\}) \qquad (77)$$

Since the above argument holds for all $i \in A$, $\theta''_A \vee \theta'''_A \in \tilde{\Theta}_A(\theta_{N\backslash A})$. This concludes the proof of Lemma 8.

First we prove the "if" direction. We do this by constructing G (and the corresponding strategy profiles). f_t is specified implicitly.

Fix, for each i, the partition points $\{\theta_i^k\}_{k=1}^K$. Initialize $k^0 := (1, 1, \ldots, 1)$, where k_i^0 denotes the ith element of this vector. Each agent i chooses whether to stay in the auction, at price $\theta_i^{k_i^0}$. i quits iff i has type $\theta_i^{k_i^0}$. Set A^0 to be the agents that do not quit. (These are the active bidders.) Set $S^0 := \emptyset$. (These are the satisfied bidders.)

At each stage, we define $\theta_{N\backslash A^{l-1}}^Q \equiv \{\theta_i^{k_i^{l-1}}\}_{i\in N\backslash A^{l-1}}$. These are the recorded type (intervals) of the agents who are no longer active.

For $l = 1, 2, \ldots$:

1. If $A^{l-1} = \emptyset$, then terminate the algorithm at allocation $y = S^{l-1}$.

2. If
$$(\theta_i^{k_i^{l-1}})_{i\in A^{l-1}} = \sup\{\tilde{\Theta}_{A^{l-1}}(\theta_{N\backslash A^{l-1}}^Q)\} \qquad (78)$$

   then

   (a) Choose agent $i \in A^{l-1}$ such that, if $\theta_{A^{l-1}} > (\theta_j^{k_j^{l-1}})_{j\in A^{l-1}}$, then $i \in f_y(\theta_{A^{l-1}}, \theta_{N\backslash A^{l-1}}^Q)$.

47

(b) Charge that agent the price $\theta_i^{k_i^{l-1}}$.

(c) Ask that agent to report $\hat{k} > k_i^{l-1}$ such that $\theta_i \in (\theta_i^{\hat{k}-1}, \theta_i^{\hat{k}}]$. Set $(k_j^l)_{j \in N}$ such that:

$$k_j^l := \begin{cases} \hat{k} & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \tag{79}$$

(d) Set $A^l := A^{l-1} \setminus i$

(e) Set $S^l := S^{l-1} \cup i$

(f) Skip to stage $l + 1$.

3. Choose $i \in A^{l-1}$ such that $(k_j^l)_{j \in N}$ satisfies

$$k_j^l := \begin{cases} k_j^{l-1} + 1 & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \tag{80}$$

and

$$(\theta_j^{k_j^l})_{j \in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q) \tag{81}$$

4. Offer agent i the option to quit. Agent i quits iff his type is less than or equal to $\theta_i^{k_i^l}$.

5. If agent i does not quit, set $A^l := A^{l-1}$.

6. If agent i quits, set $A^l := A^{l-1} \setminus i$.

7. Set $S^l := S^{l-1}$

8. Go to stage $l + 1$.

The above algorithm defines an auction with monotonically ascending prices, where an agent has the option to quit (for a transfer of zero) whenever her price rises. i makes a payment equal to her going price at the first point where she 'clinches the object' - i.e. when she is guaranteed to be in the final allocation. When i clinches the object, i is also asked to report her type - this may affect the payoffs of the agents that remain active, but does not affect her.[28]

---

[28]We could remove this feature by restricting attention to non-bossy allocation rules, where if changing i's type changes the allocation, then it also changes whether i is satisfied. However, the canonical results for SP do not assume non-bossiness of f_y, and we do not do so here. Alternatively, we could rule out such OSP mechanisms by instead requiring full implementation. However, the canonical monotonicity results for SP hold only for weak implementation.

By inspection, it is an obviously dominant strategy for i to stay in the auction if the price is less than her type, to quit if the price is greater than or equal to her type and to report her type truthfully at the point when she clinches the object.

It remains to show that the algorithm is well-defined for all type profiles, and that whenever it terminates, the resulting allocation agrees with f_y. In particular, we must show that in Steps (2a) and (3), we can pick agent i satisfying the requirements of the algorithm.

Step (2a) is well-defined by assumption.

Lemma 9. *Under the above algorithm, for all l, $(\theta_i^{k_i^l})_{i \in A^l} \in \tilde{\Theta}_{A^l}(\theta_{N \setminus A^l}^Q)$*

We prove Lemma 9 by induction. It hold for $l = 0$ by the assumption that for all i, for all θ_{-i}, $i \notin f_y(\underline{\theta}_i, \theta_{-i})$. Suppose it holds for $l - 1$. We now prove that it holds for l (assuming, of course, that the algorithm does not terminate in Step 1 of iteration l).

Suppose $(\theta_i^{k_i^{l-1}})_{i \in A^{l-1}} = \sup\{\tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q)\}$, so that Step 2 of the algorithm is triggered in iteration l. Since $(\theta_i^{k_i^{l-1}})_{i \in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q)$, we know that for all $j \in A^{l-1}$:

$$
\begin{aligned}
(\theta_i^{k_i^{l-1}})_{i \in A^{l-1}} \in \text{closure}(\{\theta_{A^{l-1}} : \forall \theta'_{A^{l-1} \setminus j} \geq (\theta_i^{k_i^{l-1}})_{i \in A^{l-1} \setminus j} : \\
j \notin f_y(\theta_j^{k_j^{l-1}}, \theta'_{A^{l-1} \setminus j}, \theta_{N \setminus A^{l-1}}^Q)\})
\end{aligned}
\tag{82}
$$

The set on the RHS of Equation 82 is upward closed with respect to the product order on $\theta_{A^{l-1} \setminus j}$. $A^l \subset A^{l-1}$ and $(\theta_i^{k_i^l})_{i \in A^l} = (\theta_i^{k_i^{l-1}})_{i \in A^l}$. Moreover, for the agent i who just clinched the object, $\theta_i^{k_i^l} > \theta_i^{k_i^{l-1}}$. Consequently, for all $j \in A^l$

$$
\begin{aligned}
(\theta_i^{k_i^l})_{i \in A^l} \in \text{closure}(\{\theta_{A^l} : \forall \theta'_{A^l \setminus j} \geq (\theta_i^{k_i^l})_{i \in A^l \setminus j} : \\
j \notin f_y(\theta_j^{k_j^l}, \theta'_{A^l \setminus j}, \theta_{N \setminus A^l}^Q)\})
\end{aligned}
\tag{83}
$$

Since this holds for each set in the intersection that defines $\tilde{\Theta}_{A^l}(\theta_{N \setminus A^l}^Q)$, this entails that $(\theta_i^{k_i^l})_{i \in A^l} \in \tilde{\Theta}_{A^l}(\theta_{N \setminus A^l}^Q)$.

Suppose $(\theta_i^{k_i^{l-1}})_{i \in A^{l-1}} \neq \sup\{\tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q)\}$, so that we reach Step 3 of the algorithm in iteration l. Then, provided Step 3 is well-defined (i.e. we can pick i satisfying our requirements),

$$(\theta_j^{k_j^l})_{j \in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q) \tag{84}$$

$A^l \subseteq A^{l-1}$, and if $i \in N \setminus A^l$, then $\theta_i^Q = \theta_i^{k_i^l}$. Thus,

$$(\theta_j^{k_j^l})_{j \in A^l} \in \tilde{\Theta}_{A^l}(\theta_{N \setminus A^l}^Q) \tag{85}$$

So we need only show that Step 3 is well-defined for iteration l, given that Lemma 9 holds for $l-1$. This will simultaneously prove Lemma 9, and demonstrate that Step 3 is well-defined throughout.

We know that

$$(\theta_j^{k_j^{l-1}})_{j \in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q) \tag{86}$$

and

$$(\theta_j^{k_j^{l-1}})_{j \in A^{l-1}} \neq \sup\{\tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q)\} \tag{87}$$

Let $\hat{A}$ be the set of all agents in A^{l-1} such that $\theta_j^{k_j^{l-1}}$ is less than the jth element of $\sup\{\tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q)\}$.

Now, we define $(\theta_j')_{j \in \hat{A}}$. For each $j \in \hat{A}$, $\theta_j' = (\theta_j^{k_j^{l-1}} + \theta_j^{k_j^{l-1}+1})/2$. Now we define two disjoint open sets:

$$\Theta_{A^{l-1}}^L = \{\theta_{A^{l-1}} : \theta_{\hat{A}} < (\theta_j')_{j \in \hat{A}}\} \tag{88}$$

$$\Theta_{A^{l-1}}^H = [\text{closure}(\Theta_{A^{l-1}}^L)]^C \tag{89}$$

The sets $\tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q) \cap \Theta_{A^{l-1}}^L$ and $\tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q) \cap \Theta_{A^{l-1}}^H$ are disjoint nonempty sets, and are open in the subspace topology. By connectedness, there exists some $\theta_{A^{l-1}}'' \in \tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q) \setminus (\Theta_{A^{l-1}}^L \cup \Theta_{A^{l-1}}^H)$. Fix some $\theta_{A^{l-1}}''$. This has at least one dimension $i \in \hat{A}$ such that $\theta_i'' = (\theta_i^{k_i^{l-1}} + \theta_i^{k_i^{l-1}+1})/2$.

Define $\theta_{A^{l-1}}''' = (\theta_j^{k_j^{l-1}})_{j \in A^{l-1}} \vee \theta_{A^{l-1}}''$. By Lemma 8, $\tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q)$ is a join-semilattice. Thus, $\theta_{A^{l-1}}''' \in \tilde{\Theta}_{A^{l-1}}(\theta_{N \setminus A^{l-1}}^Q)$.

By construction,

$$(\theta_j^{k_j^{l-1}})_{j\in\hat{A}} \leq \theta_{\hat{A}}''' < (\theta_j^{k_j^{l-1}+1})_{j\in\hat{A}} \tag{90}$$

$$\theta_{A^{l-1}\backslash\hat{A}}''' = (\theta_j^{k_j^{l-1}})_{j\in(A^{l-1}\backslash\hat{A})} \tag{91}$$

Moreover, $\theta_{A^{l-1}}'''$ has at least one dimension $i \in \hat{A}$ such that $\theta_i''' = (\theta_i^{k_i^{l-1}} + \theta_i^{k_i^{l-1}+1})/2$. Since f_y admits a finite partition and $\tilde{\Theta}_{A^{l-1}}(\theta_{N\backslash A^{l-1}}^Q)$ is closed, it follows that for

$$k_j^l := \begin{cases} k_j^{l-1} + 1 & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \tag{92}$$

$$(\theta_j^{k_j^l})_{j\in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N\backslash A^{l-1}}^Q) \tag{93}$$

This proves Lemma 9.

Now we show that whenever the algorithm terminates, it agrees with f_y.

By the assumption that for all i, for all θ_{-i}, $i \notin f_y(\underline{\theta}_i, \theta_{-i})$, it follows that all the agents that quit at price $\theta_i^1 = \underline{\theta}_i$ are never satisfied (i.e. $i \notin f_y(\theta^*)$ for the true type profile θ^*).

By construction, after any iteration l, the bidders that remain active A^l have true types $(\theta_i^*)_{i\in A^l}$ that strictly exceed the going prices $(\theta_i^{k_i^l})_{i\in A^l}$. The bidders that are inactive have their types recorded (as accurately as we need given the finite partition) in the vector $\theta_{N\backslash A^l}^Q$.

Suppose Step 1 is not activated and Step 2 is activated, in iteration l. Then, based on the information revealed up to iteration $l-1$, we know that the chosen bidder i is such that $i \in f_y(\theta^*)$. Thus, for all l, $S^l \subseteq f_y(\theta^*)$.

Suppose neither Step 1 nor Step 2 is activated in iteration l. By Lemma 9, $(\theta_j^{k_j^{l-1}})_{j\in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N\backslash A^{l-1}}^Q)$. Consider the chosen bidder i whose price is incremented. The new price vector satisfies:

$$k_j^l := \begin{cases} k_j^{l-1} + 1 & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \tag{94}$$

and

$$(\theta_j^{k_j^l})_{j\in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N\backslash A^{l-1}}^Q) \tag{95}$$

Define $(\theta_j'')_{j\in A^{l-1}} \equiv .5(\theta_j^{k_j^{l-1}})_{j\in A^{l-1}} + .5(\theta_j^{k_j^l})_{j\in A^{l-1}}$.

By $(\theta_j^{k_j^l})_{j\in A^{l-1}} \in \tilde{\Theta}_{A^{l-1}}(\theta_{N\setminus A^{l-1}}^Q)$,

$$
(\theta_j^{k_j^l})_{j\in A^{l-1}} \in
$$
$$
\text{closure}(\{\theta_{A^{l-1}} : \forall \theta'_{A^{l-1}\setminus i} \geq \theta_{A^{l-1}\setminus i} : i \notin f_y(\theta_i, \theta'_{A^{l-1}\setminus i}, \theta_{N\setminus A^{l-1}})\}) \qquad (96)
$$

The set on the RHS is (by f_y monotone) downward-closed with respect to Θ_i. Thus,

$$
(\theta''_j)_{j\in A^{l-1}} \in
$$
$$
\text{closure}(\{\theta_{A^{l-1}} : \forall \theta'_{A^{l-1}\setminus i} \geq \theta_{A^{l-1}\setminus i} : i \notin f_y(\theta_i, \theta'_{A^{l-1}\setminus i}, \theta_{N\setminus A^{l-1}})\}) \qquad (97)
$$

Thus, we can choose $(\theta'''_j)_{j\in A^{l-1}} \in \{\theta_{A^{l-1}} : \forall \theta'_{A^{l-1}\setminus i} \geq \theta_{A^{l-1}\setminus i} : i \notin f_y(\theta_i, \theta'_{A^{l-1}\setminus i}, \theta_{N\setminus A^{l-1}}^Q)\}$ such that $|(\theta'''_j)_{j\in A^{l-1}} - (\theta''_j)_{j\in A^{l-1}}| < \epsilon$, where ϵ is strictly less than half of the length of the smallest interval in the finite partition.

$\{\theta_{A^{l-1}} : \forall \theta'_{A^{l-1}\setminus i} \geq \theta_{A^{l-1}\setminus i} : i \notin f_y(\theta_i, \theta'_{A^{l-1}\setminus i}, \theta_{N\setminus A^{l-1}}^Q)\}$ is upward closed with respect to the product order on $\Theta_{A^{l-1}\setminus i}$. Thus, from the properties of $(\theta'''_j)_{j\in A^{l-1}}$, and the assumption that f_y admits a finite partition, we conclude that, for all $\theta_i \in (\theta_i^{k_i^{l-1}}, \theta_i^{k_i^l}]$, for all $\theta'_{A^{l-1}\setminus i} > (\theta_j^{k_j^{l-1}})_{j\in A^{l-1}\setminus i}$, $i \notin f_y(\theta_i, \theta'_{A^{l-1}\setminus i}, \theta_{N\setminus A^{l-1}}^Q)$.

Thus, whenever some bidder i's going price rises in iteration l, the types who quit are those that, based on the information revealed so far, are required by the allocation rule not to be satisfied. For all l, for the true type profile θ^*, $(A^l)^C \cap (S^l)^C \subseteq f_y(\theta^*)^C$.

Gathering results: For all l, $S^l = (A^l)^C \cap S^l \subseteq f_y(\theta^*)$ and $(A^t)^C \cap (S^l)^C \subseteq f_y(\theta^*)^C$. Thus, whenever $A^l = \emptyset$, $f_y(\theta^*) = S^l$. This completes the proof of the "if" direction.

Now the "only if" direction. G OSP-implements (f_y, f_t), so f_y is SP-implementable. Thus, f_y is monotone.

For all i, type $\underline{\theta}_i$ is never satisfied, and always has a zero transfer. Thus, by Theorem 3, we can restrict our attention to monotone price mechanisms that satisfy the "Either" clause in Definition 15 - i.e. every agent faces an ascending price associated with being satisfied, and a fixed outside option (call this an ascending price mechanism, or APM). Suppose we have some G that OSP-implements (f_y, f_t). Moreover, suppose G is pruned, so that G is an APM.

Take any $A \subseteq N$ and $\theta_{N \setminus A}$. We now show that $\tilde{\Theta}_A(\theta_{N \setminus A})$ is connected. Let $p : [0, 1] \to \Theta_A$ be the price path under G faced by agents in A, when the type profile for the agents in A is $\sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$ and the type profile for the agents in $N \setminus A$ is $\theta_{N \setminus A}$. Let z be the terminal history that results from that type profile, and let l be the number of elements of that sequence. Let $h_1, h_2, \ldots, h_l$ be the subhistories of z. (If z is infinitely long, instead let l be the index of some finite history such that all agents have only singleton action sets afterwards.)

Formally, p is defined as follows: Start $f(0) = (\underline{\theta}_i)_{i \in A}$. For each subhistory h_m, let $p(\frac{m}{l})$ be equal to the prices faced by agents in A at h_m. For all points in $r \in (\frac{m-1}{l}, \frac{m}{l})$, $p(r) = (1 - \beta)p(m - 1) + \beta p(m)$, for $\beta = (r - \frac{m-1}{l})/(1/l)$.

By inspection, p is a continuous function. Moreover, since at any point when an agent i quits under G, $i \notin f_y(\theta^*)$ based on the information revealed so far, for all r, $p(r) \in \tilde{\Theta}_A(\theta_{N \setminus A})$. Thus, p is a path from $\tilde{\Theta}_A(\theta_{N \setminus A})$ to $\sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$.

By Lemma 8, $\tilde{\Theta}_A(\theta_{N \setminus A})$ is a join semi-lattice. We can generate a path p' from any $\theta'_A \in \tilde{\Theta}_A(\theta_{N \setminus A})$ to $\sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$, by defining $p'(r) \equiv \theta'_A \vee p(r)$. Thus, $\tilde{\Theta}_A(\theta_{N \setminus A})$ is path-connected, which implies that it is connected.

We now show that there exists $i \in A$ such that, if $\theta_A > \sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$, then $i \in f_y(\theta_A, \theta_{N \setminus A})$. If we cannot choose some $\theta_A > \sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$, then this holds vacuously. Thus, fix some $\theta'_A > \sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$.

Let z be the terminal history in G when the type profile is $(\theta'_A, \theta_{N \setminus A})$.

Suppose there does not exist $i \in A$ such that such that, for all $\theta_A > \sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$, $i \in f_y(\theta_A, \theta_{N \setminus A})$. By definition of $\sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$, there also does not exist $i \in A$ such that such that, for all $\theta_A > \sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$, $i \notin f_y(\theta_A, \theta_{N \setminus A})$.

Thus, the price path for agents in A along history z (defined as before) is not such that $p(r) \leq \sup\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$ for all $r \in [0, 1]$. Thus, there must be a first point along z where the price path is not in $\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$. Consider the agent i whose price was incremented at that point.

For all $j \in A$, the relevant set in the intersection that defines $\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$ is upward-closed with respect to the product order on $A \setminus j$. Thus, when the price first leaves $\{\tilde{\Theta}_A(\theta_{N \setminus A})\}$ at some subhistory h_t, it must be that

$$p(\frac{t}{l}) \notin \text{closure}(\{\theta_A : \forall \theta'_{A \setminus i} \geq \theta_{A \setminus i} : i \notin f_y(\theta_i, \theta'_{A \setminus i}, \theta_{N \setminus A})\}) \qquad (98)$$

The complement of the set on the RHS of Equation 98 is open. Thus, for some $\epsilon > 0$, an open ϵ-ball around $p(\frac{t}{l})$ is a subset of the complement of

the RHS.

Consequently, we can choose some θ_i'' strictly greater than i's old price, and strictly less than i's new going price, and some $\theta_{A\setminus i}''$ strictly greater (in the product order) that the going prices for $A\setminus i$, such that $i \in f_y(\theta_i'', \theta_{A\setminus i}'', \theta_{N\setminus A})$. Since G is an APM, the actions of types $(\theta_i'', \theta_{A\setminus i}'')$ and the actions of types θ_A' are indistinguishable prior to that point. Thus, G does not result in the prescribed outcome for type profile $(\theta_i'', \theta_{A\setminus i}'', \theta_{N\setminus A})$, a contradiction.

This completes the proof of the "only-if" direction.

$\square$