

This work is distributed as a Discussion Paper by the  
**STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH**



SIEPR Discussion Paper No. 16-015

## Obviously Strategy-Proof Mechanisms

By

Shengwu Li

Stanford Institute for Economic Policy Research  
Stanford University  
Stanford, CA 94305  
(650) 725-1874

The Stanford Institute for Economic Policy Research at Stanford University supports research bearing on economic and public policy issues. The SIEPR Discussion Paper Series reports on research and policy analysis conducted by researchers affiliated with the Institute. Working papers in this series reflect the views of the authors and not necessarily those of the Stanford Institute for Economic Policy Research or Stanford University

# Obviously Strategy-Proof Mechanisms\*

Shengwu Li<sup>†</sup>

Job Market Paper

<http://ssrn.com/abstract=2560028>

First uploaded: 3 February 2015. This version: 16 March 2016.

## Abstract

What makes some strategy-proof mechanisms easier to understand than others? To address this question, I propose a new solution concept: A mechanism is *obviously strategy-proof* (OSP) if it has an equilibrium in *obviously dominant* strategies. This has a behavioral interpretation: A strategy is obviously dominant if and only if a cognitively limited agent can recognize it as weakly dominant. It also has a classical interpretation: A choice rule is OSP-implementable if and only if it can be carried out by a social planner under a particular regime of partial commitment. I fully characterize the set of OSP mechanisms in a canonical setting, with one-dimensional types and transfers. A laboratory experiment tests and corroborates the theory.

## 1 Introduction

Dominant-strategy 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.<sup>1</sup> They protect agents from strategic errors.<sup>2</sup> Dominant-strategy mechanisms prevent waste from rent-seeking espionage, since spying on other players yields no strategic advantage. Moreover, the resulting outcome does not depend sensitively on each agent’s higher-order beliefs.<sup>3</sup>

These benefits largely depend on agents *understanding* that the mechanism has an equilibrium in dominant strategies; *i.e.* that it is strategy-proof (SP). Only then can they conclude that they need not attempt to discover their opponents’ strategies or to game the system.<sup>4</sup>

---

\*I thank especially my advisors, Paul Milgrom and Muriel Niederle. I thank Nick Arnosti, Douglas Bernheim, Gabriel Carroll, Paul J. Healy, Matthew Jackson, Fuhito Kojima, Roger Myerson, Michael Ostrovsky, Matthew Rabin, Alvin Roth, and Ilya Segal for their invaluable advice. I thank Paul J. Healy for his generosity in allowing my use of the Ohio State University Experimental Economics Laboratory. I thank Muriel Niederle and the Stanford Economics Department for financial support for the experiment. This work was supported by the Kohlhaugen Fellowship Fund, through a grant to the Stanford Institute for Economic Policy Research. All errors remain my own.

<sup>†</sup>shengwu@stanford.edu

<sup>1</sup>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.”

<sup>2</sup>For instance, school choice mechanisms that lack dominant strategies may harm parents who do not strategize well (Pathak and Sönmez, 2008).

<sup>3</sup>Wilson (1987) writes, “Game Theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one player’s probability assessment about another’s preferences or information.”

<sup>4</sup>Policymakers could announce that a mechanism is strategy-proof, but that may not be enough. If agents do not understand the mechanism well, then they may be justifiably skeptical of such declarations. For instance, Google’s

However, some strategy-proof mechanisms are simpler for real people to understand than others. For instance, choosing when to quit in an ascending clock auction is the same as choosing a bid in a second-price sealed-bid auction (Vickrey, 1961). The two formats are strategically equivalent; they have the same reduced normal form.<sup>5</sup> Nonetheless, laboratory subjects are substantially more likely to play the 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 approach invokes no new primitives. Thus, it identifies a set of mechanisms as simple to understand, while remaining as parsimonious as standard game theory.

A strategy  $S_i$  is *obviously dominant* if, for any deviating strategy  $S'_i$ , starting from any earliest information set where  $S_i$  and  $S'_i$  diverge, the best possible outcome from  $S'_i$  is no better than the worst possible outcome from  $S_i$ . A mechanism is *obviously strategy-proof* (OSP) if it has an equilibrium in obviously dominant strategies. By construction, OSP depends on the extensive game form, so two games with the same normal form may differ on this criterion. Obvious dominance implies weak dominance, so OSP implies SP.

This definition distinguishes ascending auctions and second-price sealed-bid auctions. Ascending auctions are obviously strategy-proof. Suppose you value the object at \$10. If the current price is below \$10, then the best possible outcome from quitting now 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 strategy-proof, but not obviously strategy-proof. Consider the strategies “bid \$10” and “bid \$11”. The earliest information set where these diverge is the point where you submit your bid. If you bid \$11, you might win the object at some price strictly below \$10. If you bid \$10, you 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)

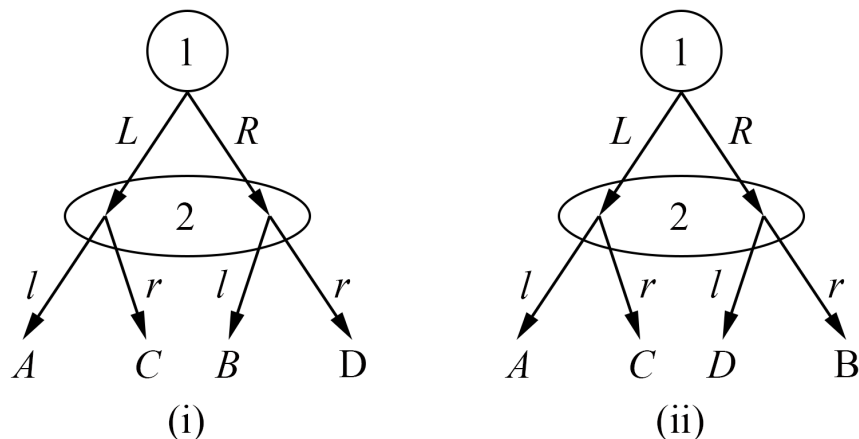
I produce two characterization theorems, which suggest two interpretations of obvious strategy-proofness. The first interpretation is behavioral: Obviously dominant strategies are those that can be recognized as dominant by a cognitively limited agent. The second interpretation is classical: Obviously strategy-proof mechanisms are those that can be carried out by a social planner with only partial commitment power.

---

advertising materials for the Generalized Second-Price auction appeared to imply that it was strategy-proof, when in fact it was not (Edelman et al., 2007). Moreover, Rees-Jones (2015) and Hassidim et al. (2015) find evidence of strategic mistakes in approximately strategy-proof matching markets, even though participants face a high-stakes decision with expert advice.

<sup>5</sup>This equivalence assumes that we restrict attention to cut-off strategies in ascending auctions.

Figure 1: ‘Similar’ mechanisms from 1’s perspective.



First, I model an agent who has a simplified mental representation of the world: Instead of understanding every detail of every game, his understanding is limited by a coarse partition on the space of all games. 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 mechanism (i), it is a weakly dominant strategy for 1 to play  $L$ . Both mechanisms are intuitively similar, but it is not a weakly dominant strategy for Agent 1 to play  $L$  in mechanism (ii).

In order for Agent 1 to recognize that it is weakly dominant to play  $L$  in mechanism (i), he must use contingent reasoning. That is, he must think through hypothetical scenarios: “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.” Notice that the quoted inferences are valid in (i), but not valid in (ii).

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, state-by-state, the outcomes after playing  $L$  are related to the outcomes after playing  $R$ . Then it is as though he cannot distinguish (i) and (ii).

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.<sup>6</sup> Two mechanisms  $G$  and  $G'$  are  *$i$ -indistinguishable* 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  $i$  the image (under the bijection) of that experience.
2. An experience might result in some outcome in  $G$  if and only if its image might result in that same outcome in  $G'$ .

With this relation, we can partition the set of all mechanisms into equivalence classes. For instance, the mechanisms in Figure 1 are 1-indistinguishable.

<sup>6</sup>An experience is a standard concept in the theory of extensive games; experiences are sometimes used to define perfect recall.

Our agent knows the experiences that a mechanism might generate, and the resulting outcomes. That is, he knows all the points at which he may be called to play, and all the actions available at each point. He knows, for any sequence of points he was called to play and actions that he took, whether the game might end and, if so, what outcomes might result. However, this does not nail down every detail of the mechanism. In particular, our agent is unable to reason contingently about hypothetical scenarios.

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 is  $i$ -indistinguishable from  $G$ .

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  $i$ -indistinguishable mechanism. In that sense, such a strategy is robustly dominant.

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, we assume that the Planner can commit to the function from all bid profiles to allocations and payments, even though each agent only directly observes his own bid. Sometimes this assumption is too strong. 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. In these settings, individual bidders may find it difficult and costly to verify the output of the auctioneer’s algorithm (Milgrom and Segal, 2015).

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 promises to Agent  $i$  must be measurable with respect to  $i$ ’s observations in the game. That is, if the Planner plays a strategy not in that subset, then there exists some agent strategy profile such that Agent  $i$  detects (with certainty) that the Planner has deviated. We call this a *bilateral commitment*.

Suppose we require that each agent  $i$ ’s strategy be optimal, for any strategies of the other agents, and for any Planner strategy compatible with the commitment made to  $i$ . 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. Consequently, 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 concept to several mechanism design environments.

For the first application, I consider binary allocation problems. In this environment, there is a set of agents  $N$  with continuous single-dimensional types  $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$ . 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 (Spence, 1974; Mirrlees, 1971; Myerson, 1981). If  $f_y$  is SP-implementable, then the required transfer rule  $f_t$  is essentially unique (Green and Laffont, 1977; Holmström, 1979).

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 environment?

I prove the following theorem: Every mechanism that OSP-implements an allocation rule is ‘essentially’ a *monotone price mechanism*, which is a new generalization of ascending clock auctions. Moreover, this is a full characterization of OSP mechanisms: For any monotone price mechanism, there exists some allocation rule that it OSP-implements.

These results imply that when we desire OSP-implementation in a binary allocation problem, we need not search the space of all extensive game forms. Without loss of generality, we can focus our attention on the class of monotone price mechanisms.<sup>7</sup>

Additionally, 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. I provide a necessary and sufficient condition for an allocation rule to be OSP-implementable.

As a second 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).

I conduct a laboratory experiment to test the theory. In the experiment, I compare three pairs of mechanisms. In each pair, both mechanisms implement the same allocation rule. One mechanism is obviously strategy-proof. The other mechanism is strategy-proof, but not obviously strategy-proof. Standard theory predicts that both mechanisms result in dominant strategy play, and have identical outcomes. Instead, subjects play the dominant strategy at significantly higher rates under the OSP mechanism, compared to the mechanism that is just SP. This effect occurs for all three pairs of mechanisms, and persists even after playing each mechanism five times with feedback.

The rest of the paper proceeds in the usual order. Section 2 reviews the literature. Section 3 provides formal definitions and characterizations. Section 4 covers applications. Section 5 reports the laboratory experiment. Section 6 concludes. Proofs omitted from the main text are in Appendix A.

## 2 Related Literature

It is widely acknowledged that ascending auctions are simpler for real bidders than second-price sealed-bid auctions (Ausubel, 2004). Laboratory experiments have investigated and corroborated

---

<sup>7</sup>Of course, if we do not impose the additional structure of a binary allocation problem, then there exist OSP mechanisms that are not monotone price mechanisms. This paper contains several examples.

this claim (Kagel et al., 1987; Kagel and Levin, 1993). More generally, laboratory subjects find it difficult to reason state-by-state about hypothetical scenarios (Charness and Levin, 2009; Esponda and Vespa, 2014; Ngangoue and Weizsacker, 2015). This mental process, often called “contingent reasoning”, has received little formal treatment in economic theory.

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.

Eyster and Rabin (2005) and Esponda (2008) model agents who do not fully account for other agents’ private information. An extensive literature on level- $k$  reasoning<sup>8</sup> models agents who hold non-equilibrium beliefs about other agents’ strategies. These are conceptually distinct from mistakes in contingent reasoning. In particular, these models predict no deviations from dominant-strategy play in strategy-proof mechanisms.<sup>9</sup>

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%).

This paper 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. To surmount this challenge, Milgrom and Segal (2015) study a class of algorithms with desirable properties, called direct deferred-acceptance threshold auctions. Their results imply that every such algorithm is OSP-implementable.

In combinatorial auction problems, finding the optimal solution is NP-hard, so the Vickrey-Clarke-Groves mechanism may be computationally infeasible. Consequently, there has been substantial interest in ‘posted-price’ mechanisms that approximate the optimum in polynomial time (Bartal et al., 2003; Feldman et al., 2014). These have the (previously unmodeled) advantage of being obviously strategy-proof.

For some mechanisms, there exist polynomial-time algorithms that verify that the mechanism is strategy-proof (Brânzei and Procaccia, 2015; Barthe et al., 2015). These are useful if agents do not trust that the mechanism is strategy-proof, but are otherwise computationally sophisticated.

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  $S_i$  *overwhelms*  $S'_i$  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. O-solvability is too strong for our current purposes, because almost no games

<sup>8</sup>Stahl and Wilson (1994, 1995); Nagel (1995); Camerer et al. (2004); Crawford and Iriberry (2007a,b).

<sup>9</sup>Level-0 agents may deviate from dominant strategy play in a strategy-proof mechanism. However, the behavior of level-0 agents is a primitive of the theory, and a sufficiently large population of level-0 agents can explain any data.

studied in mechanism design are O-solvable.<sup>10</sup>

In subsequent work, [Ashlagi and Gonczarowski \(2015\)](#) investigate the OSP-implementation of the deferred acceptance algorithm ([Gale and Shapley, 1962](#)). [Loertscher and Marx \(2015\)](#) consider the problem of bilateral trade with many buyers and many sellers, and produce a prior-free asymptotically-optimal OSP mechanism.

### 3 Definition and Characterization

The planner operates in an *environment* consisting of:

1. A set of agents,  $N \equiv \{1, \dots, n\}$ .
2. A set of outcomes,  $X$ .
3. A set of type profiles,  $\Theta \equiv \prod_{i \in N} \Theta_i$ .
4. A utility function for each agent,  $u_i : X \times \Theta_i \rightarrow \mathbb{R}$

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

1.  $H$  is a set of histories, along with a binary relation  $\prec$  on  $H$  that represents precedence.
  - (a)  $\prec$  is a partial order, and  $(H, \prec)$  form an arborescence.
  - (b)  $h_\emptyset$  denotes  $h \in H : \neg \exists h' : h' \prec h$
  - (c)  $H$  has finite depth, i.e.:

$$\exists k \in \mathbb{N} : \forall h \in H : |\{h' \in H : h' \prec h\}| \leq k \quad (2)$$

- (d)  $Z \equiv \{h \in H : \neg \exists h' : h \prec h'\}$
  - (e)  $\sigma(h)$  denotes the set of immediate successors of  $h$ .
2.  $A$  is a set of actions.
3.  $\mathcal{A} : H \setminus h_\emptyset \rightarrow A$  labels each non-initial history with the last action taken to reach it.
  - (a)  $\mathcal{A}$  is one-to-one on  $\sigma(h)$ .
  - (b)  $A(h)$  denotes the actions available at  $h$ .

$$A(h) \equiv \bigcup_{h' \in \sigma(h)} \mathcal{A}(h') \quad (3)$$

4.  $P$  is a player function.  $P : H \setminus Z \rightarrow N \cup c$
5.  $\delta_c$  is the chance function. It specifies a probability measure over chance moves.  $d_c$  denotes some realization of chance moves: For any  $h$  where  $P(h) = c$ ,  $d_c(h) \in A(h)$ .<sup>11</sup>

<sup>10</sup>For instance, neither ascending clock auctions nor second-price sealed-bid auctions are O-solvable.

<sup>11</sup>We could make the addition assumption that  $\delta_c$  has full support on the available moves  $A(h)$  when it is called to play. This ensures a pleasing invariance property: It rules out games with zero-probability chance moves that do not affect play, but do affect whether a strategy is obviously dominant. However, a full support assumption is not necessary for any of the results that follow, so we do not make it here.



6.  $\mathcal{I}_i$  is a partition of  $\{h : P(h) = i\}$  such that:
  - (a)  $A(h) = A(h')$  whenever  $h$  and  $h'$  are in the same cell of the partition.
  - (b) For any  $I_i \in \mathcal{I}_i$ , we denote:  $P(I_i) \equiv P(h)$  for any  $h \in I_i$ .  $A(I_i) \equiv A(h)$  for any  $h \in I_i$ .
  - (c) Each action is available at only one information set: If  $a \in A(I_i)$ ,  $a' \in A(I'_j)$ ,  $I_i \neq I'_j$  then  $a \neq a'$ .
7.  $g$  is an outcome function. It associates each terminal history with an outcome.  $g : Z \rightarrow X$

Additionally we denote  $I_i \prec I'_i$  if there exist  $h, h'$  such that:

1.  $h \prec h'$
2.  $h \in I_i$
3.  $h' \in I'_i$

We denote  $I_i \prec h$  if there exists  $h' \in I_i$  such that  $h' \prec h$ .  $h \prec I_i$  is defined symmetrically. We use  $\preceq$  to denote the corresponding weak order.

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  $\theta_i$ .

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  $(S, \delta_c)$ .  $z^G(h, S, d_c)$  is the result of one realization of the chance moves under  $\delta_c$ . We sometimes write this as  $z^G(h, S_i, S_{-i}, d_c)$ .

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  $\theta_i$ .

**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: Initialize  $t = 1$ ,  $h_1 = h_\emptyset$ ,  $\psi_i = \{\emptyset\}$ .

1. If  $t > 1$  and  $P(h_{t-1}) = i$ , add  $\mathcal{A}(h_t)$  to the end of  $\psi_i$ .
2. If  $P(h_t) = i$ , add  $I_i : h_t \in I_i$  to the end of  $\psi_i$ .
3. Terminate if  $h_t = h$ .
4. Set  $h_{t+1} := h' \in H : h' \in \sigma(h_t)$  and  $h' \preceq h$ .
5. Set  $t := t + 1$ .
6. Go to 1.

We use  $\Psi_i$  to denote the set  $\{\psi_i(h) : h \in H\} \cup \psi_\emptyset$ , where  $\psi_\emptyset$  is the empty sequence.<sup>12</sup>

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.

<sup>12</sup>Mandating the inclusion of the empty sequence has the following consequence: By looking at the set  $\Psi_i$ , it is not possible to infer whether  $P(h_\emptyset) = i$ .

A *choice rule* is a function  $f : \Theta \rightarrow X$ . If we consider stochastic choice rules, then it is a function  $f : \Theta \rightarrow \Delta X$ .<sup>13</sup>

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

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

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

2.  $((S_i^{\theta_i})_{\theta_i \in \Theta_i})_{i \in N}$

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

1.  $(S_i^{\theta_i})_{i \in N} \in \mathcal{C}(G, \theta)$ .

2.  $f(\theta) = g(z^G(\emptyset, (S_i^{\theta_i})_{i \in N}, \delta_c))$

Notice that each agent's strategy depends just on his own type. To ease notation, we abbreviate  $(S_i^{\theta_i})_{i \in N} \equiv S^\theta$  and  $((S_i^{\theta_i})_{\theta_i \in \Theta_i})_{i \in N} \equiv (S^\theta)_{\theta \in \Theta}$ .

Our concern is with weak implementation: We require that  $S^\theta \in \mathcal{C}(G, \theta)$ , not  $\{S^\theta\} = \mathcal{C}(G, \theta)$ . 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 “ $(G, (S^\theta)_{\theta \in \Theta})$   $\mathcal{C}$ -implements  $f$ ” to mean that  $(G, (S^\theta)_{\theta \in \Theta})$  fulfils the requirements of Definition 3. We use “ $G$   $\mathcal{C}$ -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.

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

$$\forall S'_i : \forall S'_{-i} : \mathbb{E}_{\delta_c}[u_i^G(h_\emptyset, S'_i, S'_{-i}, d_c, \theta_i)] \leq \mathbb{E}_{\delta_c}[u_i^G(h_\emptyset, S_i, S'_{-i}, d_c, \theta_i)] \quad (4)$$

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.

**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 exist  $h \in I_i, S_{-i}, d_c$  such that  $h \prec z^G(h_\emptyset, S_i, S_{-i}, d_c)$ .

3. There exist  $h \in I_i, S_{-i}, d_c$  such that  $h \prec z^G(h_\emptyset, S'_i, S_{-i}, d_c)$ .

This definition can be extended to deal with mixed strategies<sup>14</sup>, but pure strategies suffice for our current purposes.

<sup>13</sup>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.

<sup>14</sup>Three modifications are necessary: First, we change requirement 1 to be that both strategies specify different *probability measures* at  $I_i$ . Second, we adapt requirements 2 and 3 to hold for *some realization* of the mixed strategies. Finally, we include the recursive requirement, “There does not exist  $I'_i \prec I_i$  such that  $I'_i \in \alpha(S_i, S'_i)$ .”

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

$$\forall S'_i : \forall I_i \in \alpha(S_i, S'_i) : \quad (5)$$

$$\sup_{h \in I_i, S'_{-i}, d_c} u_i^G(h, S'_i, S'_{-i}, d_c, \theta_i) \leq \inf_{h \in I_i, S'_{-i}, d_c} u_i^G(h, S_i, S'_{-i}, d_c, \theta_i)$$

In words,  $S_i$  is obviously dominant if, for any deviation  $S'_i$ , conditional on reaching any earliest point of departure, the best possible outcome under  $S'_i$  is no better than the worst possible outcome under  $S_i$ .<sup>15</sup>

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.<sup>16</sup> Switching to a direct revelation mechanism may not preserve obvious dominance, so the standard revelation principle does not apply.

Weak dominance treats chance moves and other players asymmetrically. Suppose  $G$  has a Bayes-Nash equilibrium, and we replace all the agents in  $N \setminus 1$  with chance moves drawn from the Bayes-Nash equilibrium distribution. Then agent 1 has a weakly dominant strategy. By contrast, obvious dominance treats chance moves and other players symmetrically.<sup>17</sup>

**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.

A mechanism is weakly group-strategy-proof if there does not exist a coalition that could deviate and all be *strictly* better off *ex post*.

**Definition 9** (Weakly Group-Strategy-Proof).  $S \in \mathbf{WGSP}(G, \theta)$  if there does not exist a coalition  $\hat{N} \subseteq N$ , deviating strategies  $\hat{S}_{\hat{N}}$ , non-coalition strategies  $S'_{N \setminus \hat{N}}$  and chance moves  $d_c$  such that: For all  $i \in \hat{N}$ :

$$u_i^G(h_\emptyset, \hat{S}_{\hat{N}}, S'_{N \setminus \hat{N}}, d_c, \theta_i) > u_i^G(h_\emptyset, S_{\hat{N}}, S'_{N \setminus \hat{N}}, d_c, \theta_i) \quad (6)$$

Obviously strategy-proofness implies weak group-strategy-proofness.

**Proposition 1.** *If  $S \in \mathbf{OSP}(G, \theta)$ , then  $S \in \mathbf{WGSP}(G, \theta)$ .*

<sup>15</sup>Obvious dominance is related to conditional dominance (Shimoji and Watson, 1998), in that both concepts evaluate a dominance relationship conditional on reaching some information set. Obvious dominance specifies that this evaluation is carried out at any earliest point of departure, using a sup-inf comparison.

<sup>16</sup>Two extensive games have the same reduced normal form if and only if they can be made identical using a small set of elementary transformations (Thompson, 1952; Elmes and Reny, 1994). Which of these transformations does not preserve obvious dominance? Elmes and Reny (1994) propose three such transformations, INT, COA, and ADD, which preserve perfect recall. Brief inspection reveals that obvious dominance is invariant under INT and COA, but varies under ADD.

<sup>17</sup>If we replace other agents with chance moves in a second-price sealed-bid auction, the remaining agent's strategy is still not obviously dominant. This prediction is consistent with the systematic mistakes that laboratory subjects make in the Becker-DeGroot-Marschak mechanism (Becker et al., 1964; Cason and Plott, 2014).

*Proof.* We prove the contrapositive. Suppose  $S \notin \mathbf{WGSP}(G, \theta)$ . Then there is a coalition  $\hat{N}$  that could jointly deviate to strategies  $\hat{S}_{\hat{N}}$  and all be strictly better off. Fix  $S'_{N \setminus \hat{N}}$  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 agent  $i$  strictly gains from that deviation,  $S \notin \mathbf{OSP}(G, \theta)$ .<sup>18</sup>  $\square$

**Corollary 1.** *If  $S \in \mathbf{OSP}(G, \theta)$ , then  $S \in \mathbf{SP}(G, \theta)$ .*

Proposition 1 suggests a question: Is a choice rule OSP-implementable if and only if it is WGSP-implementable? Proposition 6 in Subsection 4.3 shows that this is not so.

### 3.1 Cognitive limitations

In what sense is obvious dominance obvious? Intuitively, to see that  $S'_i$  is weakly dominated by  $S_i$ , the agent must understand the entire function  $u_i^G$ , and check that for *all* opponent strategy profiles  $S_{-i}$ , the payoff from  $S'_i$  is no better than the payoff from  $S_i$ . By contrast, to see that  $S'_i$  is obviously dominated by  $S_i$ , the agent need only know the *range* of the functions  $u_i^G(\cdot, S'_i, \dots)$  and  $u_i^G(\cdot, S_i, \dots)$  at any earliest point of departure. Thus, obvious dominance can be recognized even if the agent has a simplified mental model of the world. We now make this point rigorously.

We define an equivalence relation between mechanisms. In words,  $G$  and  $G'$  are  $i$ -indistinguishable 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.  $\psi_i$  is an experience in  $G$  iff  $\psi_i$ 's image is an experience in  $G'$ .
2. Outcome  $x$  could follow experience  $\psi_i$  in  $G$  iff  $x$  could follow  $\psi_i$ 's image in  $G'$ .

**Definition 10.** *Take any  $G, G' \in \mathcal{G}$ , with information partitions  $\mathcal{I}_i, \mathcal{I}'_i$  and experience sets  $\Psi_i, \Psi'_i$ .  $G$  and  $G'$  are  $i$ -indistinguishable if there exists a bijection  $\lambda_{G, G'}$  from  $\mathcal{I}_i \cup A(\mathcal{I}_i)$  to<sup>19</sup>  $\mathcal{I}'_i \cup A'(\mathcal{I}'_i)$  such that:*

1.  $\psi_i \in \Psi_i$  iff  $\lambda_{G, G'}(\psi_i) \in \Psi'_i$
2.  $\exists z \in Z : g(z) = x, \psi_i(z) = \psi_i$  iff  $\exists z' \in Z' : g'(z') = x, \psi'_i(z') = \lambda_{G, G'}(\psi_i)$

where we use  $\lambda_{G, G'}(\psi_i)$  to denote  $\{\lambda_{G, G'}(\psi_i^k)\}_{k=1}^T$ , where  $T \in \mathbb{N} \cup \infty$ .

For  $G$  and  $G'$  that are  $i$ -indistinguishable, we define  $\lambda_{G, G'}(S_i)$  to be the strategy that, given information set  $I'_i$  in  $G'$ , plays  $\lambda_{G, G'}(S_i(\lambda_{G, G'}^{-1}(I'_i)))$ .

The next theorem states that obviously dominant strategies are the strategies that can be recognized as weakly dominant, by an agent who has a simplified mental model of the world.

**Theorem 1.** *For any  $i, \theta_i$ :  $S_i$  is obviously dominant in  $G$  if and only if for every  $G'$  that is  $i$ -indistinguishable from  $G$ ,  $\lambda_{G, G'}(S_i)$  is weakly dominant in  $G'$ .*

<sup>18</sup>I thank Ilya Segal for suggesting this concise proof.

<sup>19</sup>This definition entails that  $\lambda_{G, G'}$  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  $G'$ .

The “if” direction permits a constructive proof. Suppose  $S_i$  is not obviously dominant in  $G$ . We apply a general procedure to construct  $G'$  that is  $i$ -indistinguishable from  $G$ , such that  $\lambda_{G,G'}(S_i)$  is not weakly dominant. The “only if” direction proceeds as follows: Suppose there exists some  $G'$  in the equivalence class of  $G$ , where  $\lambda_{G,G'}(S_i)$  is not weakly dominant. There exists some earliest information set in  $G'$  where  $i$  could gain by deviating. We then use  $\lambda_{G,G'}^{-1}$  to locate an information set in  $G$ , and a deviation  $S'_i$ , that do not satisfy the obvious dominance inequality. Appendix A provides the details.

One interpretation of Theorem 1 is that obviously dominant strategies are those that can be recognized as dominant given only a *coarse description* of the game form.

Another interpretation of Theorem 1 is that obviously dominant strategies are robust to *local misunderstandings*. Suppose an agent could mistake any  $G$  for any other  $i$ -indistinguishable  $G'$ . For instance, in a second-price sealed-bid auction, he might (mistakenly) conjecture that bidding above his value increases his chance of winning with positive profits. Obviously dominant strategies are those that are robust to such mistakes.<sup>20</sup>

### 3.2 Supported by bilateral commitments

Suppose the following *extended* game form  $\tilde{G}$  with consequences in  $X$ : 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 arbitrarily rich message space  $M$ . At the start of the game, each agent  $i \in N$  privately observes  $\theta_i$ . 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.

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

$$\tilde{S}_i(\theta_i, (m_k, R_k, r_k)_{k=1}^{t-1}, m_t, R_t) \in R_t \quad (7)$$

We use  $\tilde{S}_i^{\theta_i}$  to denote the strategy played by type  $\theta_i$  of agent  $i$ . Again we abbreviate  $(\tilde{S}_i^{\theta_i})_{i \in N} \equiv \tilde{S}_N^\theta$  and  $((\tilde{S}_i^{\theta_i})_{\theta_i \in \Theta_i})_{i \in N} \equiv (\tilde{S}_N^\theta)_{\theta \in \Theta}$ .

$\tilde{S}_0$  denotes a pure strategy for the Planner. We require that these have finite length, which ensures that payoffs are well-defined.

**Definition 11.**  $\tilde{S}_0$  is a **pure strategy of finite length** if there exists  $k \in \mathbb{N}$  such that: For all  $\tilde{S}_N$ :  $(\tilde{S}_0, \tilde{S}_N)$  results in the Planner sending  $k$  or fewer total queries.

---

<sup>20</sup>Obvious dominance can thus be seen as a theory of game form misconceptions, of the kind called for by Cason and Plott (2014).

$\mathbb{S}_0$  denotes the set of all pure strategies of finite length.

The standard full commitment paradigm is equivalent to allowing the Planner to commit to a unique  $\tilde{S}_0 \in \mathbb{S}_0$  (or some probability measure over  $\mathbb{S}_0$ ). Instead, we assume that for each agent, the Planner can commit to a subset  $\hat{\mathbb{S}}_0^i \subseteq \mathbb{S}_0$  that is measurable with respect to that agent's observations in the game.

This is formalized as follows: Each  $(\tilde{S}_0, \tilde{S}_N)$  results in some observation  $o_i \equiv (o_i^C, o_i^X)$ , consisting of a communication sequence between the Planner and agent  $i$ ,  $o_i^C = (m_k, R_k, r_k)_{k=1}^T$  for  $T \in \mathbb{N}$ , as well as some outcome  $o_i^X \in X$ .<sup>21</sup>  $\mathcal{O}_i$  is the set of all possible observations (for agent  $i$ ). We define  $\phi_i : \mathbb{S}_0 \times \mathbb{S}_N \rightarrow \mathcal{O}_i$ , where  $\phi_i(\tilde{S}_0, \tilde{S}_N)$  is the unique observation resulting from  $(\tilde{S}_0, \tilde{S}_N)$ . Next we define, for any  $\hat{\mathbb{S}}_0 \subseteq \mathbb{S}_0$ :

$$\Phi_i(\hat{\mathbb{S}}_0) \equiv \{o_i : \exists \tilde{S}_0 \in \hat{\mathbb{S}}_0 : \exists \tilde{S}_N : o_i = \phi_i(\tilde{S}_0, \tilde{S}_N)\} \quad (8)$$

For any  $\hat{\mathcal{O}}_i \subseteq \mathcal{O}_i$ :

$$\Phi_i^{-1}(\hat{\mathcal{O}}_i) \equiv \{\tilde{S}_0 : \forall \tilde{S}_N : \phi_i(\tilde{S}_0, \tilde{S}_N) \in \hat{\mathcal{O}}_i\} \quad (9)$$

**Definition 12.**  $\hat{\mathbb{S}}_0$  is  *$i$ -measurable* if there exists  $\hat{\mathcal{O}}_i$  such that:

$$\hat{\mathbb{S}}_0 = \Phi_i^{-1}(\hat{\mathcal{O}}_i) \quad (10)$$

Intuitively, the  $i$ -measurable subsets of  $\mathbb{S}_0$  are those such that, if the Planner deviates, then there exists an agent strategy profile such that agent  $i$  detects the deviation. Formally, the  $i$ -measurable subsets of  $\mathbb{S}_0$  are the  $\sigma$ -algebra generated by  $\Phi_i$  (where we impose the discrete  $\sigma$ -algebra on  $\mathcal{O}_i$ ).

**Definition 13.** A *mixed strategy of finite length* over  $\hat{\mathbb{S}}_0$  specifies a probability measure over a subset  $\bar{\mathbb{S}}_0 \subseteq \hat{\mathbb{S}}_0$  such that: There exists  $k \in \mathbb{N}$  such that: For all  $\tilde{S}_0 \in \bar{\mathbb{S}}_0$  and all  $\tilde{S}_N$ :  $(\tilde{S}_0, \tilde{S}_N)$  results in the Planner sending  $k$  or fewer total queries.

We use  $\Delta\hat{\mathbb{S}}_0$  to denote the mixed strategies of finite length over  $\hat{\mathbb{S}}_0$ .  $\tilde{S}_0^\Delta$  denotes an element of such a set.

**Definition 14.** A choice rule  $f$  is *supported by bilateral commitments*  $(\hat{\mathbb{S}}_0^i)_{i \in N}$  if

1. For all  $i \in N$ :  $\hat{\mathbb{S}}_0^i$  is  $i$ -measurable.
2. There exist  $\tilde{S}_0^\Delta$ , and  $(\tilde{S}_N^\theta)_{\theta \in \Theta}$  such that:
  - (a) For all  $\theta$ :  $(\tilde{S}_0^\Delta, \tilde{S}_N^\theta)$  results in  $f(\theta)$ .
  - (b)  $\tilde{S}_0^\Delta \in \Delta \cap_{i \in N} \hat{\mathbb{S}}_0^i$
  - (c) For all  $i \in N$ ,  $\theta_i$ ,  $\tilde{S}'_{N \setminus i}, \tilde{S}_0^{\Delta'} \in \Delta\hat{\mathbb{S}}_0^i$ :  $\tilde{S}_i^{\theta_i}$  is a best response to  $(\tilde{S}_0^{\Delta'}, \tilde{S}'_{N \setminus i})$  given preferences  $\theta_i$ .

Requirement 2.a is that the Planner's mixed strategy and the agent's pure strategies result in the (distribution over) outcomes required by the choice rule. Requirement 2.b is that the Planner's strategy is a (possibly degenerate) mixture over pure strategies compatible with *every* bilateral commitment. 2.c is that each agent  $i$ 's assigned strategy is weakly dominant, when we consider the Planner as a player restricted to playing mixtures over strategies in  $\hat{\mathbb{S}}_0^i$ .

<sup>21</sup>The communication sequence might be empty, which we represent using  $T = 0$ .

“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 he 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  $(\hat{S}_0^i)_{i \in N}$  that support  $f$ .*

The intuition behind the proof is as follows: A bilateral commitment  $\hat{S}_0^i$  is essentially equivalent to the Planner committing to ‘run’ only games in some  $i$ -indistinguishable equivalence class of  $\mathcal{G}$ . Consequently, we can find a set of bilateral commitments that support  $f_y$  if and only if we can find some  $(G, (S^\theta)_{\theta \in \Theta})$  such that, for every  $i$ , for every  $\theta_i$ , for every  $G'$  that is  $i$ -indistinguishable from  $G$ ,  $\lambda_{G, G'}(S_i^{\theta_i})$  is weakly dominant in  $G'$ . By Theorem 1, this holds if and only if  $f_y$  is OSP-implementable. Appendix A provides the details.

### 3.3 The pruning principle

The standard revelation principle does not hold for OSP mechanisms; converting an OSP mechanism into the corresponding direct revelation mechanism may not preserve obvious dominance. However, there is a weaker principle that substantially simplifies the analysis.<sup>22</sup>

Here we define the *pruning* of a mechanism with respect to a set of strategies (one for each type of each agent). This is the new mechanism constructed by deleting all sub-trees that are not reached given *any* type profile.

**Definition 15** (Pruning). *Take any  $G = \langle H, \prec, A, \mathcal{A}, P, \delta_c, (\mathcal{I}_i)_{i \in N}, g \rangle$ , and  $(S^\theta)_{\theta \in \Theta}$ .  $\mathcal{P}(G, (S^\theta)_{\theta \in \Theta}) \equiv \langle \tilde{H}, \tilde{\prec}, \tilde{A}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{\delta}_c, (\tilde{\mathcal{I}}_i)_{i \in N}, \tilde{g} \rangle$  is the **pruning** of  $G$  with respect to  $(S^\theta)_{\theta \in \Theta}$ , constructed as follows:*

1.  $\tilde{H} = \{h \in H : \exists \theta : \exists d_c : h \preceq z^G(\emptyset, S^\theta, d_c)\}$
2. For all  $i$ , if  $I_i \in \mathcal{I}_i$  then  $(I_i \cap \tilde{H}) \in \tilde{\mathcal{I}}_i$ .<sup>23</sup>
3.  $(\tilde{\prec}, \tilde{A}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{\delta}_c, \tilde{g})$  are  $(\prec, A, \mathcal{A}, P, \delta_c, g)$  restricted to  $\tilde{H}$ .

It turns out that, if some mechanism OSP-implements a choice rule, then the pruning of that mechanism with respect to the equilibrium strategies OSP-implements that same choice rule. Thus, while we cannot restrict our attention to direct revelation mechanisms, we can restrict our attention to ‘minimal’ mechanisms, where no histories are off the path of play.

**Proposition 2** (The Pruning Principle). *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  $f$ , then  $(\tilde{G}, (\tilde{S}^\theta)_{\theta \in \Theta})$  OSP-implements  $f$ .*

## 4 Applications

### 4.1 Binary Allocation Problems

We now consider a canonical environment,  $(N, X, \Theta, (u_i)_{i \in N})$ . 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.

<sup>22</sup>For instance, the pruning principle is required to concisely state Theorem 3.

<sup>23</sup>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  $\{h_\emptyset\} \in \tilde{\mathcal{I}}_i$ .

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

$$u_i(\theta_i, y, t) = 1_{i \in y} \theta_i + t_i \quad (11)$$

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

An *allocation rule* is a function  $f_y : \Theta \rightarrow Y$ . A choice rule is thus a combination of an allocation rule and a payment rule,  $f = (f_y, f_t)$ , where  $f_t : \Theta \rightarrow \mathbb{R}^n$ . 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  $\delta_c$  and  $d_c$ .

**Definition 16.** An allocation rule  $f_y$  is  $\mathcal{C}$ -implementable if there exists  $f_t$  such that  $(f_y, f_t)$  is  $\mathcal{C}$ -implementable.  $G$   $\mathcal{C}$ -implements  $f_y$  if there exists  $f_t$  such that  $G$   $\mathcal{C}$ -implements  $(f_y, f_t)$

**Definition 17.**  $f_y$  is monotone if for all  $i$ , for all  $\theta_{-i}$ ,  $1_{i \in f_y(\theta_i, \theta_{-i})}$  is weakly increasing in  $\theta_i$ .

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).<sup>24</sup>

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

$$f_{t,i}(\theta_i, \theta_{-i}) = -1_{i \in f_y(\theta)} \inf\{\theta'_i : i \in f_y(\theta'_i, \theta_{-i})\} + r_i(\theta_{-i}) \quad (12)$$

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  $G$ ?

We now define a monotone price mechanism. Informally, a monotone price mechanism is such that, for every  $i$ , starting from any information set where  $i$  has more than one action:

1. Either:
  - (a) There is a 'going transfer' associated with being in the allocation, which falls monotonically.
  - (b) Whenever the going transfer falls,  $i$  chooses whether to keep bidding or to quit.
  - (c) If  $i$  quits, then  $i$  is not in the allocation and receives a fixed transfer
  - (d) If the game ends:
    - i. If  $i$  is in the allocation, then  $i$  receives the going transfer.

---

<sup>24</sup>These monotonicity results are for weak SP-implementation rather than full SP-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  $v_2 > v_1$ .



- ii. If  $i$  is not in the allocation, then  $i$  receives the fixed transfer.
  - (e) If  $i$  has more than one non-quitting action:
    - i. The going transfer will not fall in future.
    - ii. An action is available that guarantees  $i$  is in the allocation.
2. Or:
- (a) There is a ‘going transfer’ associated with *not being in the allocation*, which falls monotonically.
  - (b) Whenever the going transfer falls,  $i$  chooses whether to keep bidding or to quit.
  - (c) If  $i$  quits, then  $i$  is *in the allocation* and receives a fixed transfer
  - (d) If the game ends:
    - i. If  $i$  is *not in the allocation*, then  $i$  receives the going transfer.
    - ii. If  $i$  is *in the allocation*, then  $i$  receives the fixed transfer.
  - (e) If  $i$  has more than one non-quitting action:
    - i. The going transfer will not fall in future.
    - ii. An action is available that guarantees  $i$  is *not in the allocation*.

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 high and counts downwards.

The following definition restates the above formally.

**Definition 18** (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  $\tilde{t}_i^1 : \{I_i : I_i^* \in \psi_i(I_i)\} \rightarrow \mathbb{R}$ , 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  $I_i', I_i'' \in \{I_i : I_i^* \preceq I_i\}$ :*
    - i. *If  $I_i' \prec I_i''$ , 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  $I_i' \prec I_i''$  and  $\tilde{t}_i^1(I_i') > \tilde{t}_i^1(I_i'')$ , then  $|A(I_i') \setminus A^0| = 1$ .*
    - iv. *If  $|A(I_i') \setminus A^0| > 1$ , then there exists  $a \in A(I_i')$  such that: For all  $z$  such that  $a \in \psi_i(z)$ :  $i \in g_y(z)$ .*
  - (d) *For all  $z$  where  $I_i^* \prec z$ :*
    - i. *Either:  $i \notin g_y(z)$  and  $g_{t,i}(z) = t_i^0$ .*
    - ii. *Or:  $i \in g_y(z)$  and*

$$g_{t,i}(z) = \inf_{I_i: I_i^* \preceq I_i \prec z} \tilde{t}_i^1(I_i) \quad (13)$$

2. Or: There exists a real number  $t_i^1$ , a function  $\tilde{t}_i^0 : \{I_i : I_i^* \in \psi_i(I_i)\} \rightarrow \mathbb{R}$ , 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  $I_i', I_i'' \in \{I_i : I_i^* \preceq I_i\}$ :

i. If  $I_i' \prec I_i''$ , then  $\tilde{t}_i^0(I_i') \geq \tilde{t}_i^0(I_i'')$ .

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

iv. If  $|A(I_i') \setminus A^1| > 1$ , then there exists  $a \in A(I_i')$  such that: For all  $z$  such that  $a \in \psi_i(z) : i \notin g_y(z)$ .

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

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

ii. Or:  $i \notin g_y(z)$  and

$$g_{t,i}(z) = \inf_{I_i : I_i^* \preceq I_i \prec z} \tilde{t}_i^0(I_i) \quad (14)$$

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  $f_y$ , then  $\tilde{G} \equiv \mathcal{P}(G, (S^\theta)_{\theta \in \Theta})$  is a monotone price mechanism.*

Ascending auctions are characterized by a number of fine procedural details: Each agent faces a 'going price'. That price only ever rises. At each point, the agent chooses whether to keep bidding or to quit. If an agent quits, he quits irrevocably. If the agent wins, he pays the going price. By Theorem 3, obvious dominance entails many of these details, and can be seen as an answer to the question, "Why are ascending auctions so common?"

Theorem 4 shows that Theorem 3 is a full characterization.

**Theorem 4.** *If  $G$  is a monotone price mechanism, then there exists  $f_y$  such that  $G$  OSP-implements  $f_y$ .*

To characterize the set of OSP-implementable allocation rules, we define *traversable* allocation rules. We build towards this definition step by step. All vector inequalities are in the product order.<sup>25</sup>

Consider the set:

$$\{\theta : \forall \theta'_{N \setminus i} \geq \theta_{N \setminus i} : i \notin f_y(\theta_i, \theta'_{N \setminus i})\} \quad (15)$$

These are the type profiles such that, if we hold  $i$ 's type constant and (weakly) increase every other agent's type, then  $f_y$  requires that  $i$  is not in the allocation. To build intuition, suppose we are running an ascending clock auction with agent-specific clocks. No agent has quit yet, and the

<sup>25</sup>That is,  $v \geq v'$  iff for every index  $i$ ,  $v_i \geq v'_i$ . Similarly,  $v > v'$  iff for every index  $i$ ,  $v_i > v'_i$ .

current price vector is in the set defined by Equation 15. Then we can safely advance  $i$ 's clock while respecting the allocation rule.

Suppose instead that only a subset of the agents  $B \subseteq N$  are actively bidding, and we already know the types of the other agents  $\theta_{N \setminus B}$ . Then Equation 16 is a natural generalization of Equation 15.

$$\{\theta_B : \forall \theta'_{B \setminus i} \geq \theta_{B \setminus i} : i \notin f_y(\theta_i, \theta'_{B \setminus i}, \theta_{N \setminus B})\} \quad (16)$$

**Definition 19.**  $f_y$  is **traversable** if: For all  $B \subseteq N$ , for all  $\theta_{N \setminus B}$ , for

$$\tilde{\Theta}_B(\theta_{N \setminus B}) \equiv \bigcap_{i \in B} \text{closure}(\{\theta_B : \forall \theta'_{B \setminus i} \geq \theta_{B \setminus i} : i \notin f_y(\theta_i, \theta'_{B \setminus i}, \theta_{N \setminus B})\}) \quad (17)$$

1.  $\tilde{\Theta}_B(\theta_{N \setminus B})$  is connected.
2. There exists  $i \in B$  such that, if  $\theta_B > \sup\{\tilde{\Theta}_B(\theta_{N \setminus B})\}$ , then  $i \in f_y(\theta_B, \theta_{N \setminus B})$ .

The sets defined by Equation 17 are join-semilattices.<sup>26</sup> There exist traversable allocation rules that are not monotone, and *vice versa*.<sup>27</sup>

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 extensive game forms such that play proceeds in discrete steps. OSP is not defined for continuous-time auctions, although we can approximate some of them arbitrarily finely.<sup>28</sup>

**Definition 20.**  $f_y$  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.  $\underline{\theta}_i = \theta_i^1 < \theta_i^2 < \dots < \theta_i^K = \bar{\theta}_i$ .
2. For all  $\theta_i, \theta'_i$ , for all  $\theta_{-i}$ , if there does not exist  $k$  such that  $\theta_i \leq \theta_i^k < \theta'_i$ , then  $f_y(\theta_i, \theta_{-i}) = f_y(\theta'_i, \theta_{-i})$ .

Second, we assume that the lowest type of each agent is never in the allocation, and has a zero transfer. This is a substantive restriction, and rules out, for instance, the subsidized-trade mechanisms examined by Myerson and Satterthwaite (1983).

**Theorem 5.** Assume that:

1.  $f_y$  admits a finite partition.
2. For all  $i$ , for all  $\theta_{-i}$ ,  $i \notin f_y(\underline{\theta}_i, \theta_{-i})$ .

There exists  $G$  and  $f_t$  such that:

<sup>26</sup>For a proof, see Lemma 6 in Appendix A.

<sup>27</sup>Let  $|N| = 2$  and  $\Theta_1 = \Theta_2 = [0, 1]$ . The following allocation rule is traversable but not monotone:  $i \in f_y(\theta_i, \theta_{-i})$  if and only if  $\theta_i > \theta_{-i}$  and  $\theta_i \leq \frac{1}{2}$ . The following rule is monotone but not traversable:  $i \in f_y(\theta_i, \theta_{-i})$  if and only if  $\theta_i > \frac{1}{3}$  and  $\theta_{-i} \leq \frac{2}{3}$ .

<sup>28</sup>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.

1.  $G$  OSP-implements  $(f_y, f_t)$
2. For all  $i$ , for all  $\theta_{-i}$ ,  $f_{t,i}(\underline{\theta}_i, \theta_{-i}) = 0$   
if and only if  $f_y$  is monotone and traversable.

## 4.2 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.<sup>29</sup> Each position has an associated click-through rate  $\alpha_k$ , where  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_{n-1} > 0$ . 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  $\theta_i$  who receives position  $k$  and transfer  $t$  has utility:

$$u_i(k, t, \theta_i) = \alpha_k \theta_i^k + t \quad (18)$$

The marginal utility of moving to position  $k$  from position  $k'$ , for type  $\theta_i$ , is

$$m(k, k', \theta_i) \equiv \alpha_k \theta_i^k - \alpha_{k'} \theta_i^{k'} \quad (19)$$

We make the following assumptions on the type space  $\Theta$ :

A1. *Finite*:

$$|\Theta| < \infty \quad (20)$$

A2. *Higher slots are better*:

$$\forall k \leq n - 1 : \forall \theta \in \Theta : \forall i \in N : m(k, k + 1, \theta_i) \geq 0 \quad (21)$$

A3. *Single-crossing*:<sup>30</sup>

$$\begin{aligned} \forall k \leq n - 2 : \forall \theta, \theta' \in \Theta : \forall i, j \in N : \text{If } m(k, k + 1, \theta_i) > m(k, k + 1, \theta'_j), \\ \text{then } m(k + 1, k + 2, \theta_i) > m(k + 1, k + 2, \theta'_j). \end{aligned} \quad (22)$$

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  $\alpha_1 > \alpha_2 > \dots > \alpha_{n-1} > 0$ , 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:

$$-t_i = \sum_{k=i}^{n-1} m(k, k + 1, \theta_{k+1}) \quad (23)$$

[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.

<sup>29</sup>It is trivial to extend what follows to fewer than  $n - 1$  advertising positions, but doing so would add notation.

<sup>30</sup>This assumption is not identical to the single-crossing assumption in [Yenmez \(2014\)](#). For instance, Yenmez's condition permits the second inequality in Equation 22 to be weak.

**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  $l = 1, \dots, n-1$ :

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:

$$p''_{n-l} := \inf_{\theta \in \Theta, i \in N} \{m(n-l, n-l+1, \theta_i) : m(n-l, n-l+1, \theta_i) > p'_{n-l}\} \quad (24)$$

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

$$p_{n-l-1} := \inf_{\theta \in \Theta, i \in N} \{m(n-l-1, n-l, \theta_i) : m(n-l, n-l+1, \theta_i) \geq p'_{n-l}\} \quad (25)$$

$$A_{n-l-1} := A_{n-l} \setminus i \quad (26)$$

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  $\theta'$  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. □

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.

### 4.3 Top Trading Cycles

We now produce an impossibility result 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  $\theta_i \in \mathbb{R}^n$ .  $\Theta$  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  $\theta_i^k$ . There are no money transfers.

Given preferences  $\theta$  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:

$$R' = \{i_1, i_2, \dots, i_r = i_0\} \tag{27}$$

such that each agent  $i_k$  likes  $i_{k+1}$ 's good at least as much as any other good in  $R$ . Following Roth (1982), we assume that the algorithm in question has an arbitrary, fixed way of resolving ties.

**Definition 21.** *f is a top trading cycle rule if, for all  $\theta$ ,  $f(\theta)$  is equal to the output of the following algorithm:*

1. Set  $R^1 := N$
2. For  $l = 1, 2, \dots$ :
  - (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  $R^{l+1} = \emptyset$ .

The above algorithm is of economic interest, because it finds a core allocation in an economy with indivisible goods (Shapley and Scarf, 1974).

**Proposition 4.** *If f is a top trading cycle rule, then there exists G that SP-implements f. (Roth, 1982)*

**Proposition 5.** *If f is a top trading cycle rule, then there exists G that WGSP-implements f. (Bird, 1984)*

**Proposition 6.** *If f is a top trading cycle rule and  $n \geq 3$ , then there does not exist G that OSP-implements f.*

*Proof.* SP-implementability is a hereditary property of functions. That is, if  $f$  is SP-implementable given domain  $\Theta$ , 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 6, it suffices to produce a subfunction that is not OSP-implementable.

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

$$\begin{aligned} &\text{Either } B \succ_a C \succ_a A \succ_a \dots \\ &\text{or } C \succ_a B \succ_a A \succ_a \dots \end{aligned} \tag{28}$$

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 (by Proposition 2)  $G$  OSP-implements  $f' = f$  for domain  $\Theta'$ . 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.  $\square$

Proposition 6 implies that the OSP-implementable choice rules are not identical to the WGSP-implementable choice rules.

## 5 Laboratory Experiment

Are obviously strategy-proof mechanisms easier for real people to understand? The following laboratory experiment provides a straightforward test: We compare pairs of mechanisms that implement the same allocation rule. One mechanism in each pair is SP, but not OSP. The other mechanism is OSP. Standard game theory predicts that both mechanisms will produce the same outcome. We are interested in whether subjects play the dominant strategy at higher rates under OSP mechanisms.

### 5.1 Experiment Design

The experiment is an across-subjects design, comparing three pairs of games. There are four players in each game.

For the first pair, we compare the second-price auction (**2P**) and the ascending clock auction (**AC**). In both these games, subjects bid for a money prize. Subjects have induced affiliated private values; if a subject wins the prize, he earns an amount equal to the value of the prize, minus his payments from the auction. For each subject, his value for the prize is equal to a group draw plus a private adjustment. The group draw is uniformly distributed between \$10 and \$110. The private adjustment is uniformly distributed between \$0 and \$20. All money amounts in these games are in 25-cent increments. Each subject knows his own value, but not the group draw or the private adjustment.<sup>31</sup>

<sup>31</sup>We use affiliated private values for two reasons. First, in strategy-proof auctions with independent private values, incentives for truthful bidding are weak for bidders with values near the extremes. Affiliation strengthens incentives for these bidders. Second, Kagel et al. (1987) use affiliated private values, and the first part of the experiment is designed to replicate their results.

**2P** is SP, but not OSP. In **2P**, subjects submit their bids simultaneously. The highest bidder wins the prize, and makes a payment equal to the second-highest bid. Bids are constrained to be between \$0 and \$150.<sup>32</sup>

**AC** is OSP. In **AC**, the price starts at a low value (the highest \$25 increment that is below the group draw), and counts upwards, up to a maximum of \$150. Each bidder can quit at any point.<sup>33</sup> When only one bidder is left, that bidder wins the object at the current price.

Previous studies comparing second-price auctions to ascending clock auctions have small sample sizes, given that when the same subjects play a sequence of auctions, these are plainly not independent observations. Kagel et al. (1987) compare 2 groups playing second-price auctions to 2 groups playing ascending clock auctions. Harstad (2000) compares 5 groups playing second-price auctions to 3 groups playing ascending clock auctions. (The comparison is not the main goal of either experiment.) Other studies find similar results for second-price auctions (Kagel and Levin, 1993) and for ascending clock auctions (McCabe et al., 1990), but these do not directly compare the two formats with the same value distribution and the same subject pool. When we compare **2P** and **AC**, we can see this as a high-powered replication of Kagel et al. (1987), since we now observe 18 groups playing **2P** and 18 groups playing **AC**.<sup>34</sup>

For the second pair, we compare the second-price plus- $X$  auction (**2P+X**) and the ascending clock plus- $X$  auction (**AC+X**). Subjects' values are drawn as before. However, there is an additional random variable  $X$ , which is uniformly distributed between \$0 and \$3. Subjects are not told the value of  $X$  until after the auction.

**2P+X** is SP, but not OSP. In **2P+X**, subjects submit their bids simultaneously. The highest bidder wins the prize if and only if his bid exceeds the second-highest bid plus  $X$ . If the highest bidder wins the prize, then he makes a payment equal to the second-highest bid plus  $X$ . Otherwise, no agent wins the prize, and no payments are made. In this game, it is a dominant strategy to submit a bid equal to your value.

**AC+X** is OSP. In **AC+X**, the price starts at a low value (the highest \$25 increment that is below the group draw), and counts upwards. Each bidder can quit at any point.<sup>35</sup> When only one bidder is left, the price *continues to rise* for another  $X$  dollars, and then freezes. If the highest bidder keeps bidding until the price freezes, then she wins the prize at the final price. Otherwise, no agent wins the prize and no payments are made. In this game, it is an obviously dominant strategy to keep bidding if the price is strictly below your value, and quit otherwise.

Some subjects might find **2P** or **AC** familiar, since such mechanisms occur in some natural economic environments. Differences in subject behavior might be caused by different degrees of familiarity with the mechanism. **2P+X** and **AC+X** are novel mechanisms that subjects are unlikely to find familiar. **2P+X** and **AC+X** can be seen as perturbations of **2P** and **AC**; the underlying allocation rule is made more complex while preserving the SP-OSP distinction. Thus, comparing **2P+X** and **AC+X** indicates whether the distinction between SP and OSP mechanisms holds for novel and more complicated auction formats.

In the third pair of games, subjects may receive one of four common-value money prizes. The four prize values are drawn, uniformly at random and without replacement, from the set:

$$\{\$0.00, \$0.25, \$0.50, \$0.75, \$1.00, \$1.25\} \tag{29}$$

<sup>32</sup>In both **2P** and **AC**, if there is a tie for the highest bid, then no bidder wins the object.

<sup>33</sup>During the auction, each bidder observes the number of active bidders.

<sup>34</sup>I am not aware of any previous laboratory experiment that directly compares second-price and ascending clock auctions, holding constant the value distribution and subject pool, with more than five groups playing each format.

<sup>35</sup>As in **AC**, each bidder observes the number of active bidders. However, if the number of active bidders is 1 or 2, then the computer display informs bidders that the number of active bidders is "1 or 2".



	10 rounds	10 rounds	10 rounds
Treatment 1	<b>AC</b>	<b>AC+X</b>	<b>OSP-RSD</b>
Treatment 2	<b>2P</b>	<b>2P+X</b>	<b>SP-RSD</b>
Treatment 3	<b>AC</b>	<b>AC+X</b>	<b>SP-RSD</b>
Treatment 4	<b>2P</b>	<b>2P+X</b>	<b>OSP-RSD</b>

Subjects observe the values of all four prizes at the start of each game.

In a strategy-proof random serial dictatorship (**SP-RSD**), subjects are informed of their priority score, which is drawn uniformly at random from the integers 1 to 10. They then simultaneously submit ranked lists of the four prizes. Players are processed sequentially, from the highest priority score to the lowest. Ties in priority score are broken randomly. Each player is assigned the highest-ranked prize on his list, among the prizes that have not yet been assigned. It is a dominant strategy to rank the prizes in order of their money value. **SP-RSD** is SP, but not OSP.

In an obviously strategy-proof random serial dictatorship (**OSP-RSD**), subjects are informed of their priority score. Players take turns, from the highest priority score to the lowest. When a player takes his turn, he is shown the prizes that have not yet been taken, and picks one of them. It is an obviously dominant strategy to pick the available prize with the highest money value.

**SP-RSD** and **OSP-RSD** differ from the auctions in several ways. The auctions are private-value games of incomplete information, whereas **SP-RSD** and **OSP-RSD** are common-value games of complete information. In the auctions, subjects face two sources of strategic uncertainty: They are uncertain about their opponents' valuations, and they are uncertain about their opponents' strategies (a function of valuations). By contrast, in **SP-RSD** and **OSP-RSD**, subjects face no uncertainty about their opponents' valuations.

Unlike the auctions, **SP-RSD** and **OSP-RSD** are constant-sum games, such that one player's action cannot affect total player surplus. Any effect that persists in both the auctions and the serial dictatorships is difficult to explain using social preferences, since such theories typically make different predictions for constant-sum and non-constant-sum games. Thus, in comparing **SP-RSD** and **OSP-RSD**, we test whether the SP-OSP distinction has empirical support in mechanisms that are very different from auctions.

At the start of the experiment, subjects are randomly assigned into groups of four. These groups persist throughout the experiment. Consequently, each group's play can be regarded as a single independent observation in the statistical analysis.<sup>36</sup>

Each group either plays 10 rounds of **AC**, followed by 10 rounds of **AC+X**, or plays 10 rounds of **2P**, followed by 10 rounds of **2P+X**.<sup>37</sup> At the end of each round, subjects are shown the auction result, their own profit from this round, the winning bidder's profit from this round, and the bids (in order from highest to lowest). Notice that subjects have 10 rounds of experience with a standard auction, before being presented with its unusual **+X** variant. Thus, the data from **+X** auctions

<sup>36</sup>We could, alternatively, randomly rematch subjects after each round. However, in order for subjects to avoid encountering past opponents, this would require more than 50 subjects per session, which is more than can be accommodated in a typical experimental economics laboratory. Moreover, we would have to cluster standard errors at the session level, requiring many sessions with more than 50 subjects per session. This violates our budget constraint.

<sup>37</sup>If a stage game with dominant strategies is repeated finitely many times, then the resulting repeated game typically does not have a dominant strategy. The same holds for obviously dominant strategies. Consequently, in interpreting these results as informing us about dominant strategy play, we invoke an implicit narrow framing assumption. The same assumption is made for other experiments in this literature, such as Kagel et al. (1987) and Kagel and Levin (1993).

record moderately experienced bidders grappling with a new auction format.

Next, groups are re-randomized into either 10 rounds of **OSP-RSD** or 10 rounds of **SP-RSD**. At the end of each round, subjects see which prize they have obtained, and whether their priority score was the highest, or second-highest, and so on.

Table 1 summarizes the design. Subjects had printed copies of the instructions, and the experimenter read aloud the part pertaining to each 10-round segment just before that segment began. The instructions (correctly) informed subjects that their play in earlier segments would not affect the games in later segments. The instructions did not mention dominant strategies or provide recommendations for how to play, so as to prevent confounds from the experimenter demand effect. Instructions for both SP and OSP mechanisms are of similar length and similar reading levels<sup>38</sup>, and can be found in Appendix D.

In every SP mechanism, each subject had 90 seconds to make his choice. Each subject could revise his choice as many times as he desired during the 90 seconds, and only his final choice would count. For OSP mechanisms, mean time to completion was 113.0 seconds in **AC**, 121.4 seconds in **AC+X**, and 40.5 seconds in **RSD-OSP**. However, the rules of the OSP mechanisms imply that not every subject was actively choosing throughout that time.

## 5.2 Administrative details

Subjects were paid \$20 for participating, in addition to their profits or losses from every round of the experiment. On average, subjects made \$37.54, including the participation payment. Subjects who made negative net profits received just the \$20 participation payment.

I conducted the experiment at the Ohio State University Experimental Economics Laboratory in August 2015, using z-Tree (Fischbacher, 2007). I recruited subjects from the student population using an online system. I administered 16 sessions, where each session involved 1 to 3 groups. Each session lasted about 90 minutes. In total, the data include 144 subjects in 36 groups of 4 (with 9 groups in each treatment).<sup>39</sup> 54% of subjects are male, and 21% self-report as being economics majors.

## 5.3 Statistical Analysis

The data include 4 different auction formats, with 180 auctions per format, for a total of 720 auctions.<sup>40</sup>

One natural summary statistic for each auction is the difference between the second-highest bid and the second-highest value. This is, equivalently, the difference between that auction’s closing price, and the closing price that would have occurred if all bidders played the dominant strategy. Figure 2 displays histograms of the second-highest bid minus the second-highest value, for **AC** and **2P**. Figure 3 does the same for **AC+X** and **2P+X**. If all agents are playing the dominant strategy in an auction, then the histogram for that auction will be a point mass at zero.

There is a substantial difference between the empirical distributions for OSP and SP mechanisms. If we choose a random auction from the data, how likely is it to have a closing price within \$2.00 of the dominant strategy price? An auction is 31 percentage points more likely to have a closing price within \$2.00 of the dominant strategy price under **AC** (OSP) compared to **2P** (SP).

<sup>38</sup>Both sets of instructions are approximately at a fifth-grade reading level according to the Flesch-Kincaid readability test, which is a standard measure for how difficult a piece of text is to read (Kincaid et al., 1975).

<sup>39</sup>In two cases, network errors caused crashes which prevented a group from continuing in the experiment. I recruited new subjects to replace these groups.

<sup>40</sup>In 2 out of 720 auctions, computer errors prevented bidders from correctly entering their bids. We omit these 2 observations, but including them has little effect on any of the results that follow.

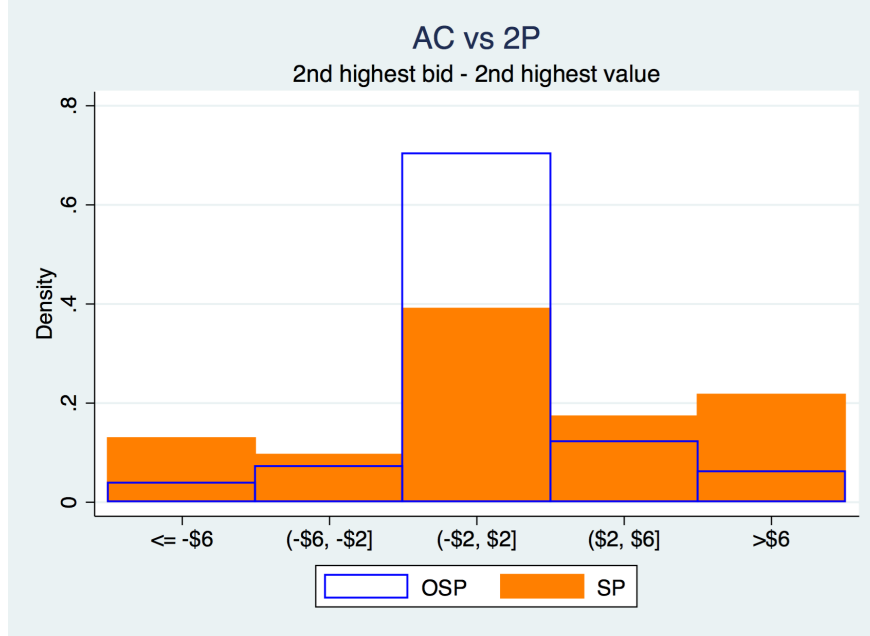


Figure 2: Histogram: 2nd-highest bid minus 2nd-highest value for **AC** and **2P**

An auction is 28 percentage points more likely to have a closing price within \$2.00 of the dominant strategy price under **AC+X** (OSP) compared to **2P+X** (SP). Closing prices under **2P+X** are systematically biased upwards ( $p = .0031$ )<sup>41</sup>.

Table 2 displays the mean absolute difference between the second-highest bid and the second-highest value, for the first 5 rounds and the last 5 rounds of each auction. This measures the magnitude of errors under each mechanism. (Alternative measures of errors are in Appendix B.) Errors are systematically larger under SP than under OSP, and this difference is significant in both the standard auctions and the novel +X auctions, and in both early and late rounds. To build intuition for effect sizes, consider that the expected profit of the winning bidder in **2P** and **AC** is about \$4.00 (given dominant strategy play). Thus, the average errors under **2P** are larger than the theoretical prediction for total bidder surplus.

There is some evidence of learning in **2P**; errors are smaller in the last five rounds compared to the first five rounds ( $p = .045$ , paired  $t$ -test). For the other three auction formats, there is no significant evidence of learning.<sup>42</sup>

To compare subject behavior under **SP-RSD** and **OSP-RSD**, we compute the proportion of games that *do not* end in the dominant strategy outcome. Under **SP-RSD**, 36.1% of games do not end in the dominant strategy outcome. Under **OSP-RSD**, 7.2% of games do not end in the dominant strategy outcome. Table 3 displays the empirical frequency of non-dominant strategy outcomes, by format and by 5-round blocks. Deviations from the dominant strategy outcome happen more frequently under **SP-RSD** than under **OSP-RSD**, and these differences are highly significant in both early and late rounds.

In **SP-RSD**, 29.0% of submitted rank-order lists contain errors. The most common error

<sup>41</sup>For each group, we take the mean difference between the second-highest bid and the second-highest value. This produces one observation per group playing **2P+X**, for a total of 18 observations, and we use a  $t$ -test for the null that these have zero mean.

<sup>42</sup> $p = .173$  for **AC**,  $p = .694$  for **2P+X**, and  $p = .290$  for **AC+X**.

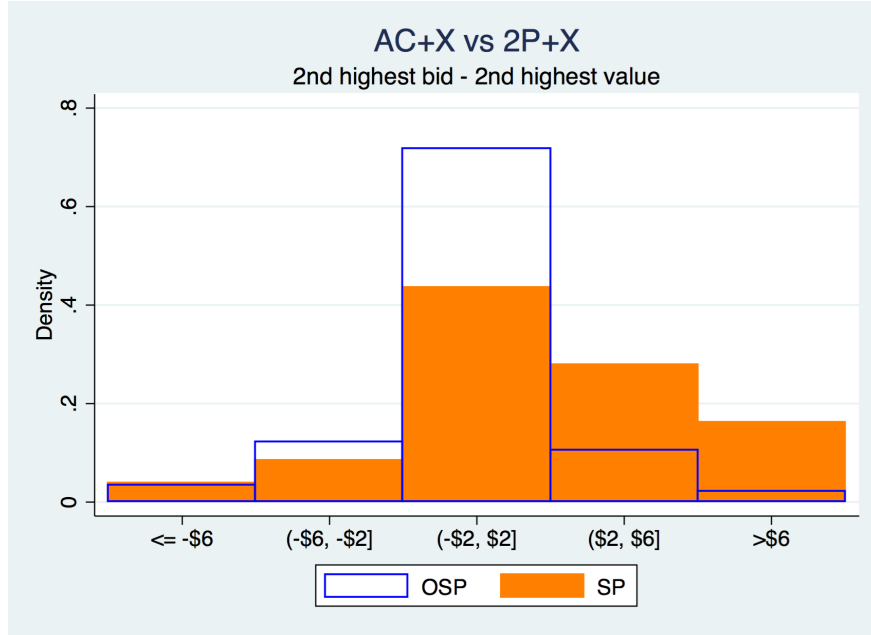


Figure 3: Histogram: 2nd-highest bid minus 2nd-highest value for **AC+X** and **2P+X**

Table 2: mean(abs(2nd bid - 2nd value))

Format	Rounds	SP	OSP	<i>p</i> -value
Auction	1-5	8.04	3.19	.006
		(1.25)	(1.05)	
	6-10	4.99	1.77	.016
		(1.18)	(0.33)	
+X Auction	1-5	3.99	1.83	.006
		(0.60)	(0.41)	
	6-10	3.69	1.29	.017
		(0.87)	(0.33)	

For each group, we take the mean absolute difference over each 5-round block. We then compute standard errors counting each group's 5-round mean as a single observation. (18 observations per cell, standard errors in parentheses.) *p*-values are computed using a two-sample *t*-test, allowing for unequal variances. Other empirical strategies yield similar results; see Appendix B for details.

Table 3: Proportion of serial dictatorships not ending in dominant strategy outcome

	SP	OSP	<i>p</i> -value
Rounds 1-5	43.3% (7.3%)	7.8% (3.3%)	.0002
Rounds 6-10	28.9% (5.2%)	6.7% (3.2%)	.0011
<i>p</i> -value	.1026	.7492	

For each group, for each 5-round block, we record the error rate. We then compute standard errors counting each group’s observed error rate as a single observation. (18 observations per cell, standard errors in parentheses.) When comparing SP to OSP, we compute *p*-values using a two-sample *t*-test, allowing for unequal variances. (Alternative empirical strategies yield similar results. See Appendix B for details.) When comparing early to late rounds of the same game, we compute *p*-values using a paired *t*-test.

under **SP-RSD** is to swap the ranks of the highest and second-highest prizes, and report the list in order 2<sup>nd</sup>-1<sup>st</sup>-3<sup>rd</sup>-4<sup>th</sup>. This accounts for 38 out of 209 incorrect rank-order lists. However, errors are diverse: No permutation of {1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>} accounts for more than a fifth of the incorrect rank-order lists.

In summary, subjects play the dominant strategy at higher rates in OSP mechanisms, as compared to SP mechanisms that should (according to standard theory) implement the same allocation rule. This difference is significant and substantial across all three pairs of mechanisms, and persists after five rounds of repeated play with feedback. Of course, the set of all strategy-proof mechanisms cannot be tested exhaustively in a single experiment. This is only preliminary evidence that obvious dominance correctly classifies strategy-proof mechanisms according to their cognitive complexity.

There are not (yet) well-formed alternative theories that explain the data presented here. The following theories predict dominant-strategy play in any strategy-proof mechanism: Level-*k* equilibrium<sup>43</sup>, cognitive hierarchy equilibrium<sup>44</sup>, cursed equilibrium<sup>45</sup>, and analogy-based expectation equilibrium<sup>46</sup>. Quantal response equilibrium predicts systematic under-bidding in **AC** and **AC+X**, which does not appear in the data.<sup>47</sup>

## 6 Discussion

In this paper, I proposed a definition of obviously strategy-proof mechanisms. One characterization theorem states: A strategy is obviously dominant if and only if a cognitively limited agent can recognize it as weakly dominant. Another characterization theorem states: A choice rule is OSP-implementable if and only if it can be supported by bilateral commitments. For binary allocation problems, I characterized the OSP mechanisms and the OSP-implementable allocation rules. I produced an impossibility result for a classic matching algorithm.

Games can be simple or complex in many ways, and no formal definition can capture all of them. In particular, since obvious dominance invokes only standard game-theoretic primitives, it does not capture framing effects. The following game has an obviously dominant strategy: You may choose any 100-digit string. You will receive \$10 if your string consists only of the digit “7”, and \$0 otherwise. This modified game also has an obviously dominant strategy: You may choose

<sup>43</sup>Stahl and Wilson (1994, 1995); Nagel (1995); Camerer et al. (2004); Crawford and Iriberri (2007a,b).

<sup>44</sup>Camerer et al. (2004).

<sup>45</sup>Eyster and Rabin (2005); Esponda (2008).

<sup>46</sup>Jehiel (2005).

<sup>47</sup>McKelvey and Palfrey (1995, 1998). See Appendix C for details.

any 100-digit string. You will receive \$10 if your string consists of the first 100 digits of  $\pi$ , and \$0 otherwise. These two games have the same extensive form, so the theory of obvious dominance cannot distinguish them.

Nonetheless, a formal standard of cognitive simplicity is valuable for several reasons. Firstly, a formal standard helps us to make simplicity an explicit design goal, by asking, “What is the optimal simple mechanism for this setting?” Secondly, a formal standard allows us to quantify trade-offs between simplicity and other design goals. For instance, one justification for using a complex mechanism is that no simple mechanism performs well for the problem at hand. Thirdly, a formal standard aids mutual understanding, since our definition of simplicity can be common knowledge, rather than relying on disparate individual intuitions.

Mechanism design typically assumes that the planner can make binding and credible promises, even about events that the promisee does not observe. Sometimes the full commitment assumption is justified, and the literature contains many excellent results for that case. However, sometimes the full commitment assumption is not justified, and we must make do with only partial commitment power. By studying OSP-implementation, we discover which standard results in mechanism design rely sensitively on the assumption of full commitment power, and learn how to design mechanisms that rely only on bilateral commitments.<sup>48</sup>

Much remains to be done. There are many classic results for SP-implementation, where OSP-implementation is an open question. For instance, in combinatorial auctions, the Vickrey-Clarke-Groves mechanism delivers first-best expected welfare but is not obviously strategy-proof (Vickrey, 1961; Clarke, 1971; Groves, 1973). However, when objects are substitutes, there always exists an obviously strategy-proof mechanism that delivers at least half of first-best expected welfare (Feldman et al., 2014).<sup>49</sup> A natural open question is: What is the welfare-maximizing obviously strategy-proof mechanism for combinatorial auctions?

## References

- Ashlagi, I. and Gonczarowski, Y. A. (2015). No stable matching mechanism is obviously strategy-proof. *arXiv preprint arXiv:1511.00452*.
- 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.
- Barthe, G., Gaboardi, M., Arias, E. J. G., Hsu, J., Roth, A., and Strub, P.-Y. (2015). Computer-aided verification in mechanism design. *arXiv preprint arXiv:1502.04052*.
- Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, 9(3):226–232.
- Bird, C. G. (1984). Group incentive compatibility in a market with indivisible goods. *Economics Letters*, 14(4):309–313.

---

<sup>48</sup>Often, full commitment power comes at a cost. If a mechanism is SP-implementable but not OSP-implementable, then its incentive properties may depend on expensive third-party audits.

<sup>49</sup>This statement is entailed by Lemma 3.4 of Feldman et al. (2014). It even holds for the more general class of fractionally subadditive preferences.

- Brânzei, S. and Procaccia, A. D. (2015). Verifiably truthful mechanisms. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 297–306. ACM.
- Camerer, C. F., Ho, T.-H., and Chong, J.-K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, pages 861–898.
- Cason, T. N. and Plott, C. R. (2014). Misconceptions and game form recognition: challenges to theories of revealed preference and framing. *Journal of Political Economy*, 122(6):1235–1270.
- Charness, G. and Levin, D. (2009). The origin of the winner’s curse: a laboratory study. *American Economic Journal: Microeconomics*, pages 207–236.
- Clarke, E. H. (1971). Multipart pricing of public goods. *Public Choice*, 11(1):17–33.
- Cramton, P. (1998). Ascending auctions. *European Economic Review*, 42(3):745–756.
- Crawford, V. P. and Iriberri, N. (2007a). Fatal attraction: Saliency, naivete, and sophistication in experimental “hide-and-seek” games. *The American Economic Review*, pages 1731–1750.
- Crawford, V. P. and Iriberri, N. (2007b). 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.
- Elmes, S. and Reny, P. J. (1994). On the strategic equivalence of extensive form games. *Journal of Economic Theory*, 62(1):1–23.
- 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.
- Feldman, M., Gravin, N., and Lucier, B. (2014). Combinatorial auctions via posted prices. *CoRR*, abs/1411.4916.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- 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.
- Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.
- Green, J. and Laffont, J.-J. (1977). Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, pages 427–438.

- Groves, T. (1973). Incentives in teams. *Econometrica: Journal of the Econometric Society*, pages 617–631.
- Harstad, R. M. (2000). Dominant strategy adoption and bidders’ experience with pricing rules. *Experimental economics*, 3(3):261–280.
- Hassidim, A., Marciano-Romm, D., Romm, A., and Shorrer, R. I. (2015). “Strategic” behavior in a strategy-proof environment. *working paper*.
- Holmström, B. (1979). Groves’ scheme on restricted domains. *Econometrica*, pages 1137–1144.
- Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic theory*, 123(2):81–104.
- 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.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., and Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, DTIC Document.
- Loertscher, S. and Marx, L. M. (2015). Prior-free bayesian optimal double-clock auctions. *working paper*.
- McCabe, K. A., Rassenti, S. J., and Smith, V. L. (1990). Auction institutional design: Theory and behavior of simultaneous multiple-unit generalizations of the Dutch and English auctions. *The American Economic Review*, pages 1276–1283.
- McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38.
- McKelvey, R. D. and Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. *Experimental economics*, 1(1):9–41.
- Milgrom, P. and Segal, I. (2015). 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.
- Myerson, R. B. and Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, pages 1313–1326.
- Ngangoue, K. and Weizsacker, G. (2015). Learning from unrealized versus realized prices. *working paper*.



- Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *The American Economic Review*, 98(4):1636–1652.
- Rees-Jones, A. (2015). Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *SSRN working paper*.
- 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.
- 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.
- Shimoji, M. and Watson, J. (1998). Conditional dominance, rationalizability, and game forms. *Journal of Economic Theory*, 83(2):161–195.
- Simon, L. K. and Stinchcombe, M. B. (1989). Extensive form games in continuous time: Pure strategies. *Econometrica*, 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.
- Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players’ models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327.
- Stahl, D. O. and Wilson, P. W. (1995). On players’ models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.
- Thompson, F. (1952). Equivalence of games in extensive form, RM 759. *RAND Corporation*.
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37.
- Wilson, R. (1987). Game theoretic analysis of trading processes. In Bewley, T., editor, *Adv. Econ. Theory*. Cambridge University Press.
- Yenmez, M. B. (2014). Pricing in position auctions and online advertising. *Economic Theory*, 55(1):243–256.

## A Proofs omitted from the main text

### A.1 Theorem 1

*Proof.* First we prove the “if” direction. Fix agent 1 and preferences  $\theta_1$ . Suppose that  $S_1$  is not obviously dominant in  $G = \langle H, \prec, A, \mathcal{A}, P, \delta_c, (\mathcal{I}_i)_{i \in N}, g \rangle$ . We need to demonstrate that there exists  $\tilde{G}$  that is  $i$ -indistinguishable from  $G$ , such that  $\lambda_{G, \tilde{G}}(S_1)$  is not weakly dominant in  $\tilde{G}$ . We proceed by construction.

Let  $(S'_1, I_1, h^{\text{sup}}, S_{-1}^{\text{sup}}, d_c^{\text{sup}}, h^{\text{inf}}, S_{-1}^{\text{inf}}, d_c^{\text{inf}})$  be such that  $I_1 \in \alpha(S_1, S'_1)$ ,  $h^{\text{inf}} \in I_1$ ,  $h^{\text{sup}} \in I_1$ , and

$$u_1^G(h^{\text{sup}}, S'_1, S_{-1}^{\text{sup}}, d_c^{\text{sup}}, \theta_1) > u_1^G(h^{\text{inf}}, S_1, S_{-1}^{\text{inf}}, d_c^{\text{inf}}, \theta_1) \quad (30)$$

Since  $G$  is a game of perfect recall, we can pick  $(S_{-1}^{\text{inf}}, d_c^{\text{inf}})$  such that  $h^{\text{inf}} \prec z^G(h_\emptyset, S_1, S_{-1}^{\text{inf}}, d_c^{\text{inf}})$ , by specifying that  $(S_{-1}^{\text{inf}}, d_c^{\text{inf}})$  plays in a way consistent with  $h^{\text{inf}}$  at any  $h \prec h^{\text{inf}}$ . Likewise for  $h^{\text{sup}}$  and  $(S_{-1}^{\text{sup}}, d_c^{\text{sup}})$ . Suppose we have so done.

We now define another game  $\tilde{G}$  is 1-indistinguishable from  $G$ . Intuitively, we construct this as follows:

1. We add a chance move at the start of the game; chance can play  $L$  or  $R$ .
2. Agent 1 does not at any history know whether chance played  $L$  or  $R$ .
3. If chance plays  $L$ , then the game proceeds as in  $G$ .
4. If chance plays  $R$ , then the game proceeds mechanically as though all players in  $N \setminus 1$  and chance played according to  $S_{-1}^{\text{sup}}, d_c^{\text{sup}}$  in  $G$ , with one exception:
5. If chance played  $R$ , we reach the information set corresponding to  $I_1$ , and agent 1 plays  $S_1(I_1)$ , then the game henceforth proceeds mechanically as though all players in  $N \setminus 1$  and chance played according to  $S_{-1}^{\text{inf}}, d_c^{\text{inf}}$  in  $G$ .

Formally, the construction proceeds as such:  $\tilde{A} = A \cup \{L, R\}$ , where  $A \cap \{L, R\} = \emptyset$ . There is a new starting history  $\tilde{h}_\emptyset$ , with two successors  $\sigma(\tilde{h}_\emptyset) = \{\tilde{h}_L, \tilde{h}_R\}$ ,  $\tilde{A}(\tilde{h}_L) = L$ ,  $\tilde{A}(\tilde{h}_R) = R$ ,  $\tilde{P}(\tilde{h}_\emptyset) = c$ . The subtree  $\tilde{H}_L \subset \tilde{H}$  starting from  $\tilde{h}_L$  ordered by  $\tilde{\prec}$  is the same as the arborescence  $(H, \prec)$ .  $(\tilde{\mathcal{A}}, \tilde{P}, \tilde{\delta}_c, \tilde{g})$  are defined on  $\tilde{H}_L$  exactly as  $(\mathcal{A}, P, \delta_c, g)$  are on  $H$ . For  $j \neq 1$ ,  $\tilde{\mathcal{I}}_j$  is defined as on  $H$ .

We now construct the subtree starting from  $\tilde{h}_R$ . Let  $h^*$  be such that  $h^* \in \sigma(h^{\text{sup}})$ ,  $h^* \preceq z^G(h^{\text{sup}}, S_1, S_{-1}^{\text{sup}}, d_c^{\text{sup}})$ .

$$\begin{aligned} H' \equiv & \{h \in H : \exists S''_1 : h \preceq z^G(h_\emptyset, S''_1, S_{-1}^{\text{sup}}, d_c^{\text{sup}})\} \\ & \cap [\{h \in H : P(h) = 1\} \cup \{h \in Z\}] \\ & \setminus \{h \in H : h^* \preceq h\} \end{aligned} \quad (31)$$

In words, these are the histories that can be reached by *some*  $S''_1$  when facing  $S_{-1}^{\text{sup}}, d_c^{\text{sup}}$ , where either agent 1 is called to play or that history is terminal, and such that those histories are not  $h^*$  or its successors.

Let  $h^{**}$  be such that  $h^{**} \in \sigma(h^{\text{inf}})$ ,  $h^{**} \preceq z^G(h^{\text{inf}}, S_1, S_{-1}^{\text{inf}}, d_c^{\text{inf}})$ .

$$\begin{aligned}
H'' \equiv & \{h \in H : \exists S_1'' : h \preceq z^G(h_\emptyset, S_1'', S_{-1}^{\text{inf}}, d_c^{\text{inf}})\} \\
& \cap [\{h \in H : P(h) = 1\} \cup \{h \in Z\}] \\
& \cap \{h \in H : h^{**} \preceq h\}
\end{aligned} \tag{32}$$

In words, these are the histories that can be reached by *some*  $S_1''$  when facing  $S_{-1}^{\text{inf}}, d_c^{\text{inf}}$ , where either agent 1 is called to play or that history is terminal, and such that those histories are  $h^{**}$  or its successors.

We now paste these together. Let  $\tilde{H}_R$  be the rooted subtree ordered by  $\tilde{\prec}$ , for some bijection  $\gamma : \tilde{H}_R \rightarrow H' \cup H''$ , such that for all  $\tilde{h}, \tilde{h}' \in \tilde{H}_R$ ,  $\tilde{h} \tilde{\prec} \tilde{h}'$  if and only if

1. EITHER:  $\gamma(\tilde{h}), \gamma(\tilde{h}') \in H'$  and  $\gamma(\tilde{h}) \prec \gamma(\tilde{h}')$
2. OR:  $\gamma(\tilde{h}), \gamma(\tilde{h}') \in H''$  and  $\gamma(\tilde{h}) \prec \gamma(\tilde{h}')$
3. OR:  $\gamma(\tilde{h}) \prec h^*$  and  $h^{**} \preceq \gamma(\tilde{h}')$

The root of this subtree exists and is unique; it corresponds to  $\gamma^{-1}(h)$ , where  $h$  is the earliest history preceding  $z^G(h_\emptyset, S_1, S_{-1}^{\text{sup}}, d_c^{\text{sup}})$  where 1 is called to play. Let  $\tilde{h}_R$  be the root of  $\tilde{H}_R$ . This completes the specification of  $\tilde{H}$ .

For all  $\tilde{h} \in \tilde{H}_R$ , we define:

1.  $\tilde{g}(\tilde{h}) = g(\gamma(\tilde{h}))$  if  $\tilde{h}$  is a terminal history.
2.  $\tilde{P}(\tilde{h}) = 1$  if  $\tilde{h}$  is not a terminal history.

For all  $\tilde{h} \in \tilde{H}_R \setminus \tilde{h}_R$ , we define  $\tilde{A}(\tilde{h}) = A(h)$ , for the unique  $(\tilde{h}', h)$  such that:

1.  $\tilde{h} \in \sigma(\tilde{h}')$
2.  $h \in \sigma(\gamma(\tilde{h}'))$
3.  $h \preceq \gamma(\tilde{h})$

We now specify the information sets for agent 1. Every  $\tilde{h} \in \tilde{H}_L$  corresponds to a unique history in  $H$ . We use  $\gamma_L$  to denote the bijection from  $\tilde{H}_L$  to  $H$ . Let  $\hat{\gamma}$  be defined as  $\gamma_L$  on  $\tilde{H}_L$  and  $\gamma$  on  $\tilde{H}_R$ .

1's information partition  $\tilde{\mathcal{I}}_1$  is defined as such:  $\forall \tilde{h}, \tilde{h}' \in \tilde{H}$  :

$$\exists \tilde{I}'_1 \in \tilde{\mathcal{I}}_1 : \tilde{h}, \tilde{h}' \in \tilde{I}'_1 \tag{33}$$

if and only if

$$\exists I'_1 \in \mathcal{I}_1 : \hat{\gamma}(\tilde{h}), \hat{\gamma}(\tilde{h}') \in I'_1 \tag{34}$$

All that remains is to define  $\delta_c$ ; we need only specify that at  $\tilde{h}_\emptyset$ ,  $c$  plays  $R$  with certainty.<sup>50</sup>

$\tilde{G} = \langle \tilde{H}, \tilde{\prec}, \tilde{A}, \tilde{\mathcal{A}}, \tilde{P}, \tilde{\delta}_c, (\tilde{\mathcal{I}}_i)_{i \in N}, \tilde{g} \rangle$  is 1-indistinguishable from  $G$ . Every experience at some history in  $\tilde{H}_L$  corresponds to some experience in  $G$ , and *vice versa*. Moreover, any experience at some history in  $\tilde{H}_R$  could also be produced by some history in  $\tilde{H}_L$ .

<sup>50</sup>If one prefers to avoid  $\delta_c$  without full support, an alternative proof puts  $\epsilon$  probability on  $L$ , for  $\epsilon$  close to 0, or for games with  $|N| > 2$  assigns  $\tilde{P}(\tilde{h}_\emptyset) = 2$ .

Let  $\lambda_{G,\tilde{G}}$  be the appropriate bijection from 1's information sets and actions in  $G$  onto 1's information sets and actions in  $\tilde{G}$ . Take arbitrary  $\tilde{S}_{-1}$ . Observe that since  $I_1 \in \alpha(S_1, S'_1)$ ,  $\lambda_{G,\tilde{G}}(S_1)$  and  $\lambda_{G,\tilde{G}}(S'_1)$  result in the same histories following  $\tilde{h}_R$ , until they reach information set  $\lambda_{G,\tilde{G}}(I_1)$ . Having reached that point,  $\lambda_{G,\tilde{G}}(S_1)$  leads to outcome  $g(z^G(h^{\text{inf}}, S_1, S_{-1}^{\text{inf}}, d_c^{\text{inf}}))$  and  $\lambda_{G,\tilde{G}}(S'_1)$  leads to outcome  $g(z^G(h^{\text{sup}}, S'_1, S_{-1}^{\text{sup}}, d_c^{\text{sup}}))$ . Thus,

$$\begin{aligned} & \mathbb{E}_{\tilde{\delta}_c} [u_1^{\tilde{G}}(\tilde{h}_\emptyset, \lambda_{G,\tilde{G}}(S'_1), \tilde{S}_{-1}, \tilde{d}_c, \theta_1)] \\ &= u_1^G(h^{\text{sup}}, S'_1, S_{-1}^{\text{sup}}, d_c^{\text{sup}}, \theta_1) \\ &> u_1^G(h^{\text{inf}}, S_1, S_{-1}^{\text{inf}}, d_c^{\text{inf}}, \theta_1) \\ &= \mathbb{E}_{\tilde{\delta}_c} [u_1^{\tilde{G}}(\tilde{h}_\emptyset, \lambda_{G,\tilde{G}}(S_1), \tilde{S}_{-1}, \tilde{d}_c, \theta_1)] \end{aligned} \quad (35)$$

So  $\lambda_{G,\tilde{G}}(S_1)$  is not weakly dominant in  $\tilde{G}$ .

We now prove the “only if” direction. Take arbitrary  $\tilde{G}$ . Suppose  $\lambda_{G,\tilde{G}}(S_1) \equiv \tilde{S}_1$  is not weakly dominant in  $\tilde{G}$ . We want to show that  $S_1$  is not obviously dominant in  $G$ .

There exist  $\theta_1, \tilde{S}'_1$  and  $\tilde{S}'_{-1}$  such that:

$$\mathbb{E}_{\tilde{\delta}_c} [u_1^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c, \theta_1)] > \mathbb{E}_{\tilde{\delta}_c} [u_1^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c, \theta_1)] \quad (36)$$

This inequality must hold for some realization of the chance function, so there exists  $\tilde{d}_c$  such that:

$$u_1^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c, \theta_1) > u_1^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c, \theta_1) \quad (37)$$

Fix  $(\tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c, \theta_1)$ .

$$z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c) \neq z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c) \quad (38)$$

Define:

$$\tilde{H}^* \equiv \{\tilde{h} \in \tilde{H} : \tilde{h} \prec z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c) \text{ and } \tilde{h} \prec z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c)\} \quad (39)$$

$$\tilde{h}^* \equiv \tilde{h} \in \tilde{H}^* : \forall \tilde{h}' \in \tilde{H}^* : \tilde{h}' \preceq \tilde{h} \quad (40)$$

Since the opponent strategies and chance moves are held constant across both sides of Equation 38,  $P(\tilde{h}^*) = 1$  and  $\tilde{h}^* \in \tilde{I}_1$ , where  $\tilde{S}_1(\tilde{I}_1) \neq \tilde{S}'_1(\tilde{I}_1)$ . Moreover,  $\tilde{I}_1 \in \alpha(\tilde{S}_1, \tilde{S}'_1)$  and  $\lambda_{\tilde{G},G}(\tilde{I}_1) \in \alpha(S_1, S'_1)$ , where we denote  $S'_1 \equiv \lambda_{\tilde{G},G}(\tilde{S}'_1)$ .

Since  $G$  and  $\tilde{G}$  are 1-indistinguishable, consider the experiences  $\lambda_{\tilde{G},G}(\psi_1(z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c)))$  and  $\lambda_{\tilde{G},G}(\psi_1(z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c)))$ .

In  $G$ ,  $\lambda_{\tilde{G},G}(\psi_1(z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c)))$  could lead to outcome  $\tilde{g}(z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c))$ . We use  $(S_{-1}^{\text{inf}}, d_c^{\text{inf}})$  to denote the corresponding opponent strategies and chance realizations that lead to that outcome. We denote  $h^{\text{inf}} \equiv h \in \lambda_{\tilde{G},G}(\tilde{I}_1) : h \prec z^G(h_\emptyset, S_1, S_{-1}^{\text{inf}}, d_c^{\text{inf}})$ .

In  $G$ ,  $\lambda_{\tilde{G},G}(\psi_1(z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c)))$  could lead to outcome  $\tilde{g}(z^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c))$ . We use  $(S_{-1}^{\text{sup}}, d_c^{\text{sup}})$  to denote the corresponding opponent strategies and chance realizations that lead to that outcome. We denote  $h^{\text{sup}} \equiv h \in \lambda_{\tilde{G},G}(\tilde{I}_1) : h \prec z^G(h_\emptyset, S'_1, S_{-1}^{\text{sup}}, d_c^{\text{sup}})$ .

$$\begin{aligned}
& u_1^G(h^{\sup}, S'_1, S_{-1}^{\sup}, d_c^{\sup}, \theta_1) \\
& = u_1^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}'_1, \tilde{S}'_{-1}, \tilde{d}_c, \theta_1) \\
& > u_1^{\tilde{G}}(\tilde{h}_\emptyset, \tilde{S}_1, \tilde{S}'_{-1}, \tilde{d}_c, \theta_1) \\
& = u_1^G(h^{\inf}, S_1, S_{-1}^{\inf}, d_c^{\inf}, \theta_1)
\end{aligned} \tag{41}$$

where  $h^{\sup}, h^{\inf} \in \lambda_{\tilde{G}, G}(\tilde{I}_1)$  and  $\lambda_{\tilde{G}, G}(\tilde{I}_1) \in \alpha(S_1, S'_1)$ . Thus  $S_1$  is not obviously dominant in  $G$ .  $\square$

## A.2 Theorem 2

*Proof.* The key is to see that, for every  $G \in \mathcal{G}$ , there is a corresponding  $\tilde{S}_0^\Delta$ , and *vice versa*. We use  $\bar{\mathbb{S}}_0$  to denote the support of  $\tilde{S}_0^\Delta$ . In particular, observe the following isomorphism:

Table 4: Equivalence between extensive game forms and Planner mixed strategies

$G$	$\tilde{S}_0^\Delta$
$d_c$	$\tilde{S}_0 \in \bar{\mathbb{S}}_0$
$\delta_c$	the probability measure specified by $\tilde{S}_0^\Delta$
$g(z)$ for $z \in Z$	the Planner's choice of outcome when she ends the game
$I_i$	$((m_k, R_k, r_k)_{k=1}^{t-1}, m_t, R_t)$ consistent with some $\tilde{S}_0 \in \bar{\mathbb{S}}_0$
$A(I_i)$	$R_t$
$\psi_i(z)$	$o_i^c$ consistent with some $\tilde{S}_0 \in \bar{\mathbb{S}}_0$ and $\tilde{S}_N$

Information sets in  $G$  are equivalent to sequences of past communication  $((m_k, R_k, r_k)_{k=1}^{t-1}, m_t, R_t)$  under  $\tilde{S}_0^\Delta$ . Available actions at some information set  $A(I_i)$  are equivalent to acceptable responses  $R_t$ . Thus, for any strategy in some game  $G$ , we can construct an equivalent strategy given appropriate  $\tilde{S}_0^\Delta$ , and *vice versa*.

Furthermore, fixing a chance realization  $d_c$  and agent strategies  $S_N$  uniquely results in some outcome. Similarly, fixing a realization of the planner's mixed strategy  $\tilde{S}_0 \in \bar{\mathbb{S}}_0$  and agent strategies  $\tilde{S}_N$  uniquely determines some outcome. Consequently, for any  $G \in \mathcal{G}$ , there exists  $\tilde{S}_0^\Delta$  with the same strategies available for each agent and the same resulting (probability measure over) outcomes, and *vice versa*.<sup>51</sup> Table 4 summarizes.

The next step is to see that a bilateral commitment  $\hat{\mathbb{S}}_0^i$  is equivalent to the Planner promising to 'run' only games in some equivalence class that is  $i$ -indistinguishable.

Suppose that there is some  $G$  that OSP-implements  $f$ . Pick some equivalent  $\tilde{S}_0^\Delta$  with support  $\bar{\mathbb{S}}_0$ . For each  $i \in N$ , specify the bilateral commitment  $\hat{\mathbb{S}}_0^i \equiv \Phi_i^{-1}(\Phi_i(\bar{\mathbb{S}}_0))$ . These bilateral commitments support  $f$ .

To see this, take any  $\tilde{S}_0^{\Delta'} \in \Delta \hat{\mathbb{S}}_0^i$ , with support  $\bar{\mathbb{S}}_0'$ . For any  $\tilde{S}'_0 \in \bar{\mathbb{S}}_0'$ , for any  $\tilde{S}'_N$ , there exists  $\tilde{S}_0 \in \bar{\mathbb{S}}_0$  and  $\tilde{S}_N$  such that  $\phi_i(\tilde{S}'_0, \tilde{S}'_N) = \phi_i(\tilde{S}_0, \tilde{S}_N)$ . By construction,  $G$  is such that: There exists  $z \in Z$  where  $\psi_i(z)$  and  $g(z)$  are equivalent to  $\phi_i(\tilde{S}_0, \tilde{S}_N)$ . Thus, for  $G'$  that is equivalent to  $\tilde{S}_0^{\Delta'}$ , every terminal history in  $G'$  results in the same experience for  $i$  and the same outcome as some terminal history in  $G$ . Consequently,  $G$  and  $G'$  are  $i$ -indistinguishable. Thus, by Theorem 1, the strategy assigned to agent  $i$  with type  $\theta_i$  is weakly dominant in  $G'$ , which implies that it is a best

<sup>51</sup>Implicitly, this relies on the requirement that both  $G$  and  $\tilde{S}_0^\Delta$  have finite length. If one had finite length but the other could be infinitely long, the resulting outcome would not be well defined and the equivalence would not hold.

response to  $\tilde{S}_0^{\Delta'}$  and any  $\tilde{S}_N$  in the bilateral commitment game. Thus, if  $f$  is OSP-implementable, then  $f$  can be supported by bilateral commitments.

Suppose that  $f$  can be supported by bilateral commitments  $(\hat{S}_0^i)_{i \in N}$ , with requisite  $\tilde{S}_0^\Delta$  (with support  $\tilde{S}_0$ ) and  $(\tilde{S}_N^\theta)_{\theta \in \Theta}$ . Without loss of generality, let us suppose these are ‘minimal’ bilateral commitments, *i.e.*  $\hat{S}_0^i = \Phi_i^{-1}(\Phi_i(\tilde{S}_0))$ . Pick  $G$  that is equivalent to  $\tilde{S}_0^\Delta$ .  $G$  OSP-implements  $f$ .

To see this, consider any  $G'$  such that  $G$  and  $G'$  are  $i$ -indistinguishable. Let  $\tilde{S}_0^{\Delta'}$  denote the Planner strategy that corresponds to  $G'$ . At any terminal history  $z'$  in  $G'$ , the resulting experience  $\psi_i(z')$  and outcome  $g'(z')$  are equivalent to the experience  $\psi_i(z)$  and outcome  $g(z)$  for some terminal history  $z$  in  $G$ . These in turn correspond to some observation  $o_i \in \Phi_i(\tilde{S}_0)$ . Thus  $\tilde{S}_0^{\Delta'} \in \Delta \hat{S}_0^i$ . Since  $f$  is supported by  $(\hat{S}_0^j)_{j \in N}$ ,  $\tilde{S}_i^{\theta_i}$  is a best response (for type  $\theta_i$ ) to  $\tilde{S}_0^{\Delta'}$  and any  $\tilde{S}_{N \setminus i}$ . Thus, the equivalent strategy  $S_i^{\theta_i}$  is weakly dominant in  $G'$ . Since this argument holds for all  $i$ -indistinguishable  $G'$ , by Theorem 1,  $S_i^{\theta_i}$  is obviously dominant in  $G$ . Thus, if  $f$  can be supported by bilateral commitments, then  $f$  is OSP-implementable.  $\square$

### A.3 Proposition 2

*Proof.* We prove the contrapositive. Suppose  $(\tilde{G}, (\tilde{S}^\theta)_{\theta \in \Theta})$  does not OSP-implement  $f$ . Then there exists some  $(i, \theta_i, \tilde{S}_i^{\theta_i}, \tilde{S}'_i, \tilde{I}_i)$  such that  $\tilde{I}_i \in \alpha(\tilde{S}_i^{\theta_i}, \tilde{S}'_i)$  and

$$u_i^{\tilde{G}}(\tilde{h}, \tilde{S}_i^{\theta_i}, \tilde{S}_{-i}, \tilde{d}_c, \theta_i) < u_i^{\tilde{G}}(\tilde{h}', \tilde{S}'_i, \tilde{S}'_{-i}, \tilde{d}'_c, \theta_i) \quad (42)$$

for some  $(\tilde{h}, \tilde{S}_{-i}, \tilde{d}_c)$  and  $(\tilde{h}', \tilde{S}'_{-i}, \tilde{d}'_c)$ .

Notice that  $\tilde{h}$  and  $\tilde{h}'$  correspond to histories  $h$  and  $h'$  in  $G$ . Moreover, we can define  $S'_i = \tilde{S}'_i$  at information sets containing histories that are shared by  $G$  and  $\tilde{G}$ , and specify  $S'_i$  arbitrarily elsewhere. We do the same for  $(\tilde{S}_{-i}, \tilde{d}_c)$  and  $(\tilde{S}'_{-i}, \tilde{d}'_c)$ , to construct  $(S_{-i}, d_c)$  and  $(S'_{-i}, d'_c)$ . But, starting from  $h$  and  $h'$  respectively, these result in the same outcomes as their partners in  $\tilde{G}$ . Thus,

$$u_i^G(h, S_i^{\theta_i}, S_{-i}, d_c, \theta_i) < u_i^G(h', S'_i, S'_{-i}, d'_c, \theta_i) \quad (43)$$

$h, h' \in I_i$ , for  $I_i \in \alpha(S_i^{\theta_i}, S'_i)$ . Thus  $(G, (S^\theta)_{\theta \in \Theta})$  does not OSP-implement  $f$ .  $\square$

### A.4 Theorem 3

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

$$\Theta_h \equiv \{\theta \in \Theta : h \preceq z^G(h_\theta, S^\theta)\} \quad (44)$$

$$\Theta_{h,i} \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_h\} \quad (45)$$

For information set  $I_i$ , we define

$$\Theta_{I_i} \equiv \cup_{h \in I_i} \Theta_h \quad (46)$$

$$\Theta_{I_i,i} \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_{I_i}\} \quad (47)$$

$$\Theta_{I_i,i}^1 \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_{I_i} \text{ and } i \in f_y(\theta_i, \theta_{-i})\} \quad (48)$$

$$\Theta_{I_i,i}^0 \equiv \{\theta_i : \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in \Theta_{I_i} \text{ and } i \notin f_y(\theta_i, \theta_{-i})\} \quad (49)$$

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

**Lemma 1.** *Suppose  $(G, (S^\theta)_{\theta \in \Theta})$  OSP-implements  $(f_y, f_t)$ , where  $G \equiv \langle H, \prec, A, \mathcal{A}, P, \delta_c, (\mathcal{I}_i)_{i \in N}, g \rangle$ . For all  $i$ , for all  $I_i \in \mathcal{I}_i$ , if:*

1.  $\theta_i < \theta'_i$
2.  $\theta_i \in \Theta_{I_i,i}^1$
3.  $\theta'_i \in \Theta_{I_i,i}^0$

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

Equivalently, for any  $I_i$ , there exists  $a_{I_i}^*$  such that for all  $\theta_i \in \Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0$ ,  $S_i^{\theta_i}(I_i) = a_{I_i}^*$ .

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

$$u_i(\theta_i, h, S_i^{\theta_i}, S_{-i}) \geq u_i(\theta_i, h', S_i^{\theta'_i}, S'_{-i}) \quad (50)$$

which implies

$$\theta_i + t_i \geq t'_i \quad (51)$$

and

$$u_i(\theta'_i, h, S_i^{\theta_i}, S_{-i}) \leq u_i(\theta'_i, h', S_i^{\theta'_i}, S'_{-i}) \quad (52)$$

which implies

$$\theta'_i + t_i \leq t'_i \quad (53)$$

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

$$\theta'_i + t_i > t'_i \quad (54)$$

a contradiction. This proves Lemma 1.

**Lemma 2.** *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  $a_{I_i}^*$ .*

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 there exists  $t_i^0$  such that:*

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

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 there exists  $t_i^1$  such that:

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

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

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

$$\forall h \in I_i : \forall S_{-i} : i \notin g_y(z^G(h, S_i^{\theta'_i}, S_{-i})) \quad (55)$$

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

$$\exists h \in I_i : \exists S_{-i} : i \notin g_y(z^G(h, S_i^{\theta''_i}, S_{-i})) \quad (56)$$

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

$$\begin{aligned} & \inf_{h \in I_i, S_{-i}} g_{t,i}(z^G(h, S_i^{\theta'_i}, S_{-i})) \geq \\ & \sup_{h \in I_i, S_{-i}} \{g_{t,i}(z^G(h, S_i^{\theta''_i}, S_{-i})) : i \notin g_y(z^G(h, S_i^{\theta''_i}, S_{-i}))\} \end{aligned} \quad (57)$$

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

$$\begin{aligned} & \inf_{h \in I_i, S_{-i}} \{g_{t,i}(z^G(h, S_i^{\theta''_i}, S_{-i})) : i \notin g_y(z^G(h, S_i^{\theta''_i}, S_{-i}))\} \\ & \geq \sup_{h \in I_i, S_{-i}} g_{t,i}(z^G(h, S_i^{\theta'_i}, S_{-i})) \end{aligned} \quad (58)$$

The RHS of Equation 57 is weakly greater than the LHS of Equation 58. The RHS of Equation 58 is weakly greater than the LHS of Equation 57. 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,  $\theta''_i$  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 2. Part 2 follows by symmetry; we omit the details since they involve only small changes to the above argument.

**Lemma 3.** 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 3 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  $\tilde{I}_i \in \{I'_i \in \mathcal{I}_i : I_i \in \psi(I'_i)\}$ ,  $S_i^{\theta'_i}(\tilde{I}_i) = S_i(\tilde{I}_i)$ . Fix that  $\theta'_i$ .

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 1), 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  $h'' \in I_i, S''_{-i}$ :

$$u_i(\theta''_i, h'', S_i^{\theta''_i}, S''_{-i}) \geq u_i(\theta'_i, h, S_i^{\theta'_i}, S_{-i}) \quad (59)$$

which entails

$$t_i^0 \geq \theta''_i + g_{t,i}(z^G(h, S_i, S_{-i})) \quad (60)$$

which entails

$$t_i^0 - \sup\{\theta_i \in \Theta_{I_i,i}^0 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\} + \epsilon \geq g_{t,i}(z^G(h, S_i, S_{-i})) \quad (61)$$

But, by hypothesis,

$$t_i^0 - \sup\{\theta_i \in \Theta_{I_i,i}^0 : S_i^{\theta_i}(I_i) \neq a_{I_i}^*\} < g_{t,i}(z^G(h, S_i, S_{-i})) \quad (62)$$

Since this argument holds for all  $\epsilon > 0$ , we can pick  $\epsilon$  small enough to create a contradiction. This proves part 1 of Lemma 3. Part 2 follows by symmetry.

**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  $|\Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0| > 1$  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. (Equivalently) 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 4 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 1, 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  $\theta'_{-i}$  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 2, 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

$$\theta'_i + g_{t,i}(z^G(h', S_i^{\theta'_i}, S_{-i}^{\theta'_{-i}})) \geq t_i^0 \quad (63)$$

By  $\theta''_i \in \Theta_{I_i,i}^0$ , there exists  $h'' \in I_i$  and  $\theta''_{-i}$  such that  $i \notin g_y(z^G(h'', S_i^{\theta''_i}, S_{-i}^{\theta''_{-i}}))$ . By Lemma 2

$$g_{t,i}(z^G(h'', S_i^{\theta''_i}, S_{-i}^{\theta''_{-i}})) = t_i^0 \quad (64)$$

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

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

$$g_{t,i}(z^G(h'', S_i^{\theta''}, S_{-i}^{\theta''-i})) \geq \theta'' + g_{t,i}(z^G(h', S_i^{\theta''''}, S_{-i}^{\theta''-i})) \quad (65)$$

$$\begin{aligned} t_i^0 &\geq \theta'' + g_{t,i}(z^G(h', S_i^{\theta''''}, S_{-i}^{\theta''-i})) \\ &> \theta'' + g_{t,i}(z^G(h', S_i^{\theta''}, S_{-i}^{\theta''-i})) \end{aligned} \quad (66)$$

which contradicts Equation 63.

Part 2 is the contrapositive of Part 1. This proves Lemma 4.

**Lemma 5.** *Suppose  $(G, (S^\theta)_{\theta \in \Theta})$  OSP-implements  $(f_y, f_t)$  and  $\mathcal{P}(G, (S^\theta)_{\theta \in \Theta}) = G$ .*

*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,  $\Theta_{I_i,i} \neq \emptyset$ .

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  $h'' \in I_i$  such that  $h'' \preceq z^G(\emptyset, S_i^{\theta''_i}, S_{-i}^{\theta''_{-i}})$ . By Equation 12, 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  $I_i \in \alpha(S_i^{\theta'''_i}, S_i^{\theta'_i})$ . If  $t_i^{0''} > t_i^0$ , then

$$u_i(\theta'_i, h', S_i^{\theta'_i}, S'_{-i}) = t_i^0 < t_i^{0''} = u_i(\theta'_i, h'', S_i^{\theta'''_i}, S_{-i}^{\theta''_{-i}}) \quad (67)$$

so  $S_i^{\theta'_i}$  is not obviously dominant for  $(i, \theta'_i)$ . If  $t_i^{0''} < t_i^0$ , then

$$u_i(\theta'''_i, h', S_i^{\theta'_i}, S'_{-i}) = t_i^0 > t_i^{0''} = u_i(\theta'''_i, h'', S_i^{\theta'''_i}, S_{-i}^{\theta''_{-i}}) \quad (68)$$

so  $S_i^{\theta'''_i}$  is not obviously dominant for  $(i, \theta'''_i)$ . By contradiction, this proves Lemma 5 for this case. A symmetric argument proves Lemma 5 for the case where  $\Theta_{I_i,i}^0 = \emptyset$ .

Note that, if Lemma 5 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  $I_i^*$ ,  $|\Theta_{I_i,i}^1 \cap \Theta_{I_i,i}^0| > 1$ . Since  $\Theta_{I_i^*,i}^1 \neq \emptyset$  and  $\Theta_{I_i^*,i}^0 \neq \emptyset$ , by Lemma 4,  $I_i^*$  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 12, 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  $h' \in I_i^*$  such that  $h' \preceq z^G(\emptyset, S_i^{\theta'_i}, S_{-i}^{\theta'_{-i}})$ .

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,

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

which entails

$$\begin{aligned} t_i^0 &\geq t_i^{0'} + 1_{i \in f_y(\theta''_i, \theta'_{-i})}(\hat{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*, i}^1\}) \\ &\geq t_i^{0'} + 1_{i \in f_y(\theta''_i, \theta'_{-i})}(-\epsilon) \\ &\geq t_i^{0'} - \epsilon \end{aligned} \quad (70)$$

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

$$u_i(\theta''_i, \hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}) \leq u_i(\theta''_i, h', S_i^{\theta''_i}, S_{-i}^{\theta'_{-i}}) \quad (71)$$

which entails

$$\begin{aligned} t_i^0 &\leq t_i^{0'} + 1_{i \in f_y(\theta''_i, \theta'_{-i})}(\theta''_i - \inf\{\theta_i \in \Theta_{I_i^*, i}^1\}) \\ &= t_i^{0'} + 1_{i \in f_y(\theta''_i, \theta'_{-i})}(\theta''_i - \sup\{\theta_i \in \Theta_{I_i^*, i}^0\}) \leq t_i^{0'} \end{aligned} \quad (72)$$

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:

$$u_i(\hat{\theta}_i, \hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}) \geq u_i(\hat{\theta}_i, h', S_i^{\theta''_i}, S_{-i}^{\theta'_{-i}}) \quad (73)$$

which entails

$$\begin{aligned} t_i^0 &\geq \hat{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*, i}^1\} + t_i^{0'} \\ &\geq t_i^{0'} - \epsilon \end{aligned} \quad (74)$$

and we can pick  $\epsilon$  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  $S_i^{\hat{\theta}_i}(I_i) \neq S_i^{\theta''_i}(I_i)$  or  $S_i^{\hat{\theta}_i}(I_i) \neq S_i^{\theta''_i}(I_i)$ . Moreover,  $f_{t,i}(\hat{\theta}_i, \theta'_i) = -1_{i \in f_y(\hat{\theta}_i, \theta'_{-i})} \inf\{\theta_i \in \Theta_{I_i^*, i}^1\} + t_i^{0'}$ .

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

$$u_i(\check{\theta}_i, h', S_i^{\check{\theta}_i}, S_{-i}^{\theta'_{-i}}) \geq u_i(\check{\theta}_i, \hat{h}, S_i^{\hat{\theta}_i}, S_{-i}^{\hat{\theta}_{-i}}) \quad (75)$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta'_{-i})}(\check{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*, i}^1\}) + t_i^{0'} \geq t_i^0 \quad (76)$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta'_{-i})} \epsilon + t_i^{0'} \geq t_i^0 \quad (77)$$

and we can pick  $\epsilon$  small enough to yield a contradiction. Suppose  $S_i^{\check{\theta}_i}(I_i) \neq S_i^{\theta''_i}(I_i)$ . By Equation 12,  $f_{t,i}(\check{\theta}_i, \theta'_{-i}) = -1_{i \in f_y(\check{\theta}_i, \theta'_{-i})} \inf\{\theta_i \in \Theta_{I_i^*, i}^1\} + t_i^{0'}$ , and  $f_{t,i}(\theta''_i, \hat{\theta}_{-i}) = -\inf\{\theta_i \in \Theta_{I_i^*, i}^1\} + t_i^0$ . OSP requires

$$u_i(\check{\theta}_i, h', S_i^{\check{\theta}_i}, S_{-i}^{\theta'_{-i}}) \geq u_i(\check{\theta}_i, \hat{h}, S_i^{\theta''_i}, S_{-i}^{\hat{\theta}_{-i}}) \quad (78)$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta'_{-i})}(\check{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*, i}^1\}) + t_i^{0'} \geq (\check{\theta}_i - \inf\{\theta_i \in \Theta_{I_i^*, i}^1\}) + t_i^0 \quad (79)$$

which entails

$$1_{i \in f_y(\check{\theta}_i, \theta'_{-i})} \epsilon + t_i^{0'} \geq \epsilon + t_i^0 \quad (80)$$

which entails

$$t_i^{0'} \geq t_i^0 \quad (81)$$

a contradiction. By the above argument, for all  $I_i$  satisfying the assumptions of Lemma 5, there is a unique transfer  $t_i^0$  for all terminal histories  $z$  passing through  $I_i$  such that  $i \notin g_y(z)$ . Equation 12 thus implies that there is a unique transfer  $t_i^1$  for all terminal histories  $z$  passing through  $I_i$  such that  $i \in g_y(z)$ . Moreover,  $t_i^1 = -\inf\{\theta_i \in \Theta_{I_i^*, i}^1\} + t_i^0$ . This proves Lemma 5.

Now to bring this all together. We leave showing parts (1.c.iv) and (2.c.iv) of Definition 18 to the last. 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 Proposition 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  $I'_i \prec I_i^*$ ,  $|A(I'_i)| = 1$ . By Lemma 1, there is a unique action  $a_{I_i^*}^*$  taken by all types in  $\Theta_{I_i^*, i}^1 \cap \Theta_{I_i^*, i}^0$ .

Either  $|\Theta_{I_i^*, i}^1 \cap \Theta_{I_i^*, i}^0| > 1$  or  $|\Theta_{I_i^*, i}^1 \cap \Theta_{I_i^*, i}^0| \leq 1$ .

If  $|\Theta_{I_i^*, i}^1 \cap \Theta_{I_i^*, i}^0| > 1$ , then by Lemma 4,  $\tilde{G}$  pruned and  $|A(I_i)| \geq 2$ ,

1. EITHER: There exists  $\theta_i \in \Theta_{I_i^*, i}^0$  such that  $S_i^{\theta_i}(I_i^*) \neq a_{I_i^*}^*$ , and for all  $\theta'_i \in \Theta_{I_i^*, i}^1$ ,  $S_i^{\theta'_i}(I_i^*) = a_{I_i^*}^*$ .
2. OR: There exists  $\theta_i \in \Theta_{I_i^*, i}^1$  such that  $S_i^{\theta_i}(I_i^*) \neq a_{I_i^*}^*$ , and for all  $\theta'_i \in \Theta_{I_i^*, i}^0$ ,  $S_i^{\theta'_i}(I_i^*) = a_{I_i^*}^*$ .

In the first case, then by Lemma 2, there is some  $t_i^0$  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  $g_{t,i}(z^G(h, S_i, S_{-i})) = t_i^0$ . Moreover, we can define a ‘going transfer’ at all information sets  $I_i^*$  such that  $I_i^* \preceq I_i'$ :

$$\tilde{t}_i^1(I_i') \equiv \min_{I_i'' : I_i^* \preceq I_i'' \preceq I_i'} [t_i^0 - \sup\{\theta_i \in \Theta_{I_i'', i}^0 : S_i^{\theta_i}(I_i'') \neq a_{I_i''}^*\}] \quad (82)$$

Notice that this function falls monotonically as we move along the game tree; for any  $I_i', I_i''$  such that  $I_i' \preceq I_i''$ ,  $\tilde{t}_i^1(I_i') \geq \tilde{t}_i^1(I_i'')$ . 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  $i \notin y$ , and by Lemma 2 this yields transfer  $t_i^0$ . We define  $A^0$  to include all such quitting actions; i.e.  $A^0$  is the set of all actions such that:

1.  $a \in I_i$  for some  $I_i \in \mathcal{I}_i$
2. For all  $z$  such that  $a \in \psi_i(z)$ :  $i \notin g_y(z)$  and  $g_{t,i}(z) = t_i^0$

Lemma 3 and SP together imply that, at any terminal history  $z$ , if  $i \in g_y(z)$ , then

$$g_{t,i}(z) = \inf_{I_i : I_i^* \preceq I_i \prec z} \tilde{t}_i^1(I_i) \quad (83)$$

This holds because, if  $g_{t,i}(z) < \inf_{I_i : I_i^* \preceq I_i \prec z} \tilde{t}_i^1(I_i)$ , then type  $\theta_i$  such that  $t_i^0 - \inf_{I_i : I_i^* \preceq I_i \prec z} \tilde{t}_i^1(I_i) < \theta_i < t_i^0 - g_{t,i}(z)$  could profitably deviate to play  $a \in A^0$  at information set  $I_i^*$ .

In the second case, then by Lemma 2, 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 transfer’ at all information sets  $I_i'$  such that  $I_i^* \preceq I_i'$ :

$$\tilde{t}_i^0(I_i') \equiv \min_{I_i'' : I_i^* \preceq I_i'' \preceq I_i'} [t_i^1 + \inf\{\theta_i \in \Theta_{I_i'', i}^0 : S_i^{\theta_i}(I_i'') \neq a_{I_i''}^*\}] \quad (84)$$

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  $i \in y$ , and transfer  $t_i^1$ .
3. For any  $z$ , if  $i \notin g_y(z)$ , then

$$g_{t,i}(z) = \inf_{I_i : I_i^* \preceq I_i \prec z} \tilde{t}_i^0(I_i) \quad (85)$$

We define  $A^1$  symmetrically for this second case.

Part (1.c.iii) and (2.c.iii) of Definition 18 follow from Lemma 5. 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 5.

Now for the last piece: We prove that parts (1.c.iv) and (2.c.iv) of Definition 18 hold. The proof of part (1.c.iv) is as follows: Suppose we are facing the ‘either’ clause of Definition 18, and for some  $I_i$ ,  $|A(I_i') \setminus A^0| > 1$ . By part (1.c.iii), we know that the going transfer  $\tilde{t}_i^1$  can fall no further. Since  $G$  is pruned and  $|A(I_i') \setminus A^0| > 1$ , there exist two distinct types of  $i$ ,  $\theta_i, \theta_i' \in \Theta_{I_i', i}$ , who do not quit at  $I_i'$ , and take different actions. Since neither quits at  $I_i'$  and the going transfer falls no further, there exist  $\theta_{-i}, \theta_{-i}' \in \Theta_{I_i', -i}$  such that  $i \in f_y(\theta_i, \theta_{-i})$  and  $i \in f_y(\theta_i', \theta_{-i}')$ . So there exist  $(h \in I_i', S_{-i})$  and  $(h' \in I_i', S_{-i}')$  such that

$$i \in g_y(h, S_i^{\theta_i}, S_{-i}) \quad (86)$$

$$i \in g_y(h', S_i^{\theta'_i}, S'_{-i}) \quad (87)$$

$$g_{t,i}(h, S_i^{\theta_i}, S_{-i}) = g_{t,i}(h', S_i^{\theta'_i}, S'_{-i}) = \tilde{t}_i^1(I'_i) \quad (88)$$

WLOG suppose  $\theta_i < \theta'_i$ . Suppose that there does not exist  $a \in A(I'_i)$  such that, for all  $z$  such that  $a \in \psi_i(z)$ ,  $i \in g_y(z)$ . Then there must exist  $(h'' \in I'_i, S''_{-i})$  such that

$$i \notin g_y(h'', S_i^{\theta'_i}, S''_{-i}) \quad (89)$$

$$g_{t,i}(h'', S_i^{\theta'_i}, S''_{-i}) = t_i^0 \quad (90)$$

Note that  $I'_i \in \alpha(S_i^{\theta_i}, S_i^{\theta'_i})$ . But then  $S_i^{\theta'_i}$  is not obviously dominant, a contradiction, since

$$\begin{aligned} u_i^{\tilde{G}}(h'', S_i^{\theta'_i}, S''_{-i}, \theta'_i) &= t_i^0 \leq \theta_i + \tilde{t}_i^1(I'_i) \\ &< \theta'_i + \tilde{t}_i^1(I'_i) = u_i^{\tilde{G}}(h, S_i^{\theta_i}, S_{-i}, \theta'_i) \end{aligned} \quad (91)$$

(The first inequality holds because of type  $\theta_i$ 's incentive constraint.) This shows that part (1.c.iv) of Definition 18 holds. Part (2.c.iv) is proved symmetrically.  $\square$

## A.5 Theorem 4

*Proof.* Take any monotone price mechanism  $G$ . For any  $i$ , the following strategy  $S_i$  is obviously dominant:

1. If  $i$  encounters an information set consistent with Clause 1 of Definition 18, then, from that point forward:
  - (a) If  $\theta_i + \tilde{t}_i^1(I_i) > t_i^0$  and there exists  $a \in A(I_i) \setminus A^0$ , play  $a \in A(I_i) \setminus A^0$ .
    - i. If  $|A(I_i) \setminus A^0| > 1$ , then play  $a \in A(I'_i)$  such that: For all  $z$  such that  $a \in \psi_i(z)$ :  $i \in g_y(z)$ .
  - (b) Else play some  $a \in A^0$ .<sup>52</sup>
2. If  $i$  encounters an information set consistent with Clause 2 of Definition 18, then, from that point forward:
  - (a) If  $\theta_i + t_i^1 < \tilde{t}_i^0(I_i)$  and there exists  $a \in A(I_i) \setminus A^1$ , play  $a \in A(I_i) \setminus A^1$ .
    - i. If  $|A(I_i) \setminus A^1| > 1$ , then play  $a \in A(I'_i)$  such that: For all  $z$  such that  $a \in \psi_i(z)$ :  $i \notin g_y(z)$ .
  - (b) Else play some  $a \in A^1$ .

---

<sup>52</sup>If  $|A^0 \cap A(I_i)| > 1$ , the agent chooses deterministically but arbitrarily.

The above strategy is well-defined for any agent in any monotone price mechanism, by inspection of Definition 18.

Consider any deviating strategy  $S'_i$ . At any earliest point of departure, the agent will have encountered an information set consistent with either Clause 1 or Clause 2 of Definition 18. Suppose that the agent has encountered an information set covered by Clause 1.

Take some earliest point of departure  $I_i \in \alpha(S_i, S'_i)$ . Notice that, by (1.d) of Definition 18, *no matter what strategy  $i$  plays*, conditional on reaching  $I_i$ , either agent  $i$  is not in the allocation and receives  $t_i^0$ , or agent  $i$  is in the allocation and receives a transfer  $\hat{t}_i \leq \tilde{t}_i^1(I_i)$ .

Suppose  $\theta_i + \tilde{t}_i^1(I_i) > t_i^0$ . Note that under  $S_i$ , conditional on reaching  $I_i$ , the agent either is not in the allocation and receives  $t_i^0$ , or is in the allocation and receives a transfer strictly above  $t_i^0 - \theta_i$ . If  $S'_i(I_i) \in A_0$  (*i.e.* if agent  $i$  quits), then the best outcome under  $S'_i$  is no better than the worst outcome under  $S_i$ . If  $S'_i(I_i) \notin A_0$ , then, since  $S'_i(I_i) \neq S_i(I_i)$ ,  $|A(I_i) \setminus A^0| > 1$ . Then, by (1.c.iii) of Definition 18,  $\tilde{t}_i^1$  will fall no further. So  $S_i(I_i)$  guarantees that  $i$  will be in the allocation and receive transfer  $\tilde{t}_i^1(I_i)$ . But, by (1.d) of Definition 18, the best possible outcome under  $S'_i$  conditional on reaching  $I_i$  is no better, so the obvious dominance inequality holds.

Suppose  $\theta_i + \tilde{t}_i^1(I_i) \leq t_i^0$ . Then, under  $S_i$ , conditional on reaching  $I_i$ , agent  $i$  is not in the allocation and has transfer  $t_i^0$ .<sup>53</sup> However, under  $S'_i$ , either the outcome is the same, or agent  $i$  is in the allocation for some transfer  $\hat{t}_i \leq \tilde{t}_i^1(I_i) \leq t_i^0 - \theta_i$ . Thus, the best possible outcome under  $S'_i$  is no better than the worst possible outcome under  $S_i$ , and the obvious dominance inequality holds.

The argument proceeds symmetrically for Clause 2.

Notice that the above strategies result in *some* allocation and *some* payments, as a function of the type profile. We define these to be  $(f_y, f_t)$ , such that  $G$  OSP-implements  $(f_y, f_t)$ . □

## A.6 Theorem 5

*Proof.*

**Lemma 6.** *For all  $B \subseteq N$ , for all  $\theta_{N \setminus B}$ ,  $\tilde{\Theta}_B(\theta_{N \setminus B})$  is a join-semilattice with respect to the product order on  $\Theta_B$ .*

Take any  $\theta''_B, \theta'''_B \in \tilde{\Theta}_B(\theta_{N \setminus B})$ . We want to show that  $\theta''_B \vee \theta'''_B \in \tilde{\Theta}_B(\theta_{N \setminus B})$ .

For all  $i \in B$ ,

$$\theta''_B, \theta'''_B \in \text{closure}(\{\theta_B : \forall \theta'_{B \setminus i} \geq \theta_{B \setminus i} : i \notin f_y(\theta_i, \theta'_{B \setminus i}, \theta_{N \setminus B})\}) \quad (92)$$

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

Consider  $\check{\theta}_B \equiv \theta''_B \vee \theta'''_B$ . Its  $i$ th element has the property:  $\check{\theta}_i = \max\{\theta''_i, \theta'''_i\}$ . WLOG, suppose  $\theta''_i \geq \theta'''_i$ . Then, since  $\check{\theta}_{B \setminus i} \geq \theta''_{B \setminus i}$ ,

$$\check{\theta}_B \in \text{closure}(\{\theta_B : \forall \theta'_{B \setminus i} \geq \theta_{B \setminus i} : i \notin f_y(\theta_i, \theta'_{B \setminus i}, \theta_{N \setminus B})\}) \quad (93)$$

Since the above argument holds for all  $i \in B$ ,  $\theta''_B \vee \theta'''_B \in \tilde{\Theta}_B(\theta_{N \setminus B})$ . This concludes the proof of Lemma 6.

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  $k^0 := (1, 1, \dots, 1)$ , where  $k_i^0$  denotes the  $i$ th element of this vector. Each agent  $i$  chooses whether to stay in the auction, at price  $\theta_i^{k_i^0}$ .  $i$

<sup>53</sup>By (1.c.ii) of Definition 18, either  $i$  will have quit in the past, or will have an opportunity to quit now, which he exercises.

quits iff  $i$  has a type weakly less than  $\theta_i^{k_i^0}$ . Set  $B^0 := N$ . (We use  $B^l$  to keep track of the agents still bidding in iteration  $l$ .) Set  $y^0 := \emptyset$ . (We use  $y^l$  to keep track of agents who must be in the allocation at iteration  $l$ .)

At each stage, we define  $\theta_{N \setminus B^{l-1}}^Q \equiv \{\theta_i^{k_i^{l-1}}\}_{i \in N \setminus B^{l-1}}$ . These are the recorded type (intervals) of the agents who have stopped bidding.

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

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

2. If

$$(\theta_i^{k_i^{l-1}})_{i \in B^{l-1}} = \sup\{\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)\} \quad (94)$$

then

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

(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:

$$k_j^l := \begin{cases} \hat{k} & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \quad (95)$$

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

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

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

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

$$k_j^l := \begin{cases} k_j^{l-1} + 1 & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \quad (96)$$

and

$$(\theta_j^{k_j^l})_{j \in B^{l-1}} \in \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q) \quad (97)$$

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  $B^l := B^{l-1}$ .

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

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

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.<sup>54</sup>

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 7.** *(Given the assumptions of Theorem 5.) Under the above algorithm, for all  $l$ ,  $(\theta_i^{k_i^l})_{i \in B^l} \in \tilde{\Theta}_{B^l}(\theta_{N \setminus B^l}^Q)$*

We prove Lemma 7 by induction. It holds for  $l = 0$  by the assumption that for all  $i$ , for all  $\theta_{-i}$ ,  $i \notin f_y(\theta_i, \theta_{-i})$ . 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  $(\theta_i^{k_i^{l-1}})_{i \in B^{l-1}} = \sup\{\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)\}$ , so that Step 2 of the algorithm is triggered in iteration  $l$ . Since  $(\theta_i^{k_i^{l-1}})_{i \in B^{l-1}} \in \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$ , we know that for all  $j \in B^{l-1}$ :

$$\begin{aligned} (\theta_i^{k_i^{l-1}})_{i \in B^{l-1}} \in \text{closure}(\{ \theta_{B^{l-1}} : \forall \theta'_{B^{l-1} \setminus j} \geq (\theta_i^{k_i^{l-1}})_{i \in B^{l-1} \setminus j} : \\ j \notin f_y(\theta_j^{k_j^{l-1}}, \theta'_{B^{l-1} \setminus j}, \theta_{N \setminus B^{l-1}}^Q) \}) \end{aligned} \quad (98)$$

The set on the RHS of Equation 98 is upward closed with respect to the product order on  $\theta_{B^{l-1} \setminus j}$ .  $B^l \subset B^{l-1}$  and  $(\theta_i^{k_i^l})_{i \in B^l} = (\theta_i^{k_i^{l-1}})_{i \in B^l}$ . Moreover, for the agent  $i$  who just clinched the object,  $\theta_i^{k_i^l} > \theta_i^{k_i^{l-1}}$ . Consequently, for all  $j \in B^l$

$$\begin{aligned} (\theta_i^{k_i^l})_{i \in B^l} \in \text{closure}(\{ \theta_{B^l} : \forall \theta'_{B^l \setminus j} \geq (\theta_i^{k_i^l})_{i \in B^l \setminus j} : \\ j \notin f_y(\theta_j^{k_j^l}, \theta'_{B^l \setminus j}, \theta_{N \setminus B^l}^Q) \}) \end{aligned} \quad (99)$$

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

Suppose  $(\theta_i^{k_i^{l-1}})_{i \in B^{l-1}} \neq \sup\{\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)\}$ , 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),

$$(\theta_j^{k_j^l})_{j \in B^{l-1}} \in \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q) \quad (100)$$

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

---

<sup>54</sup>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.

$$(\theta_j^{k_j^l})_{j \in B^l} \in \tilde{\Theta}_{B^l}(\theta_{N \setminus B^l}^Q) \quad (101)$$

So we need only show that Step 3 is well-defined for iteration  $l$ , given that Lemma 7 holds for  $l - 1$ . This will simultaneously prove Lemma 7, and demonstrate that Step 3 is well-defined throughout.

We know that

$$(\theta_j^{k_j^{l-1}})_{j \in B^{l-1}} \in \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q) \quad (102)$$

and

$$(\theta_j^{k_j^{l-1}})_{j \in B^{l-1}} \neq \sup\{\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)\} \quad (103)$$

Let  $\hat{B} \subseteq B^{l-1}$  be the set of all agents in  $B^{l-1}$  such that  $\theta_j^{k_j^{l-1}}$  is less than the  $j$ th element of  $\sup\{\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)\}$ . Define  $\theta'_{\hat{B}} = 0.5(\theta_j^{k_j^{l-1}})_{j \in B^{l-1}} + 0.5(\theta_j^{k_j^{l-1}+1})_{j \in B^{l-1}}$ .

Now we define two disjoint open sets:

$$\Theta_{B^{l-1}}^L = \{\theta_{B^{l-1}} : \theta_{\hat{B}} < \theta'_{\hat{B}}\} \quad (104)$$

$$\Theta_{B^{l-1}}^H = [\text{closure}(\Theta_{B^{l-1}}^L)]^C \quad (105)$$

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

Define  $\theta'''_{B^{l-1}} = (\theta_j^{k_j^{l-1}})_{j \in B^{l-1}} \vee \theta''_{B^{l-1}}$ . By Lemma 6,  $\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$  is a join-semilattice. Thus,  $\theta'''_{B^{l-1}} \in \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$ .

By construction,

$$(\theta_j^{k_j^{l-1}})_{j \in \hat{B}} \leq \theta'''_{\hat{B}} < (\theta_j^{k_j^{l-1}+1})_{j \in \hat{B}} \quad (106)$$

$$\theta'''_{B^{l-1} \setminus \hat{B}} = (\theta_j^{k_j^{l-1}})_{j \in (B^{l-1} \setminus \hat{B})} \quad (107)$$

Pick  $i \in \hat{B}$  such that  $\theta''_i = (\theta_i^{k_i^{l-1}} + \theta_i^{k_i^{l-1}+1})/2$ . Define:

$$k_j^l := \begin{cases} k_j^{l-1} + 1 & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \quad (108)$$

Consider the sequence  $\frac{1}{t}\theta'''_{B^{l-1}} + (1 - \frac{1}{t})(\theta_j^{k_j^l})_{j \in B^{l-1}}$  for  $t = 1, 2, 3, \dots$ . Since  $f_y$  admits a finite partition, and this sequence does not move past any partition coordinates, every element of this sequence is in  $\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$ . Since  $\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$  is closed, the limit of the sequence,  $(\theta_j^{k_j^l})_{j \in B^{l-1}}$ , is also in  $\tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$ . This proves Lemma 7.

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

By the assumption that for all  $i$ , for all  $\theta_{-i}$ ,  $i \notin f_y(\underline{\theta}_i, \theta_{-i})$ , 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  $\theta^*$ ).

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

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  $l$ ,  $y^l \subseteq f_y(\theta^*)$ .

Suppose neither Step 1 nor Step 2 is activated in iteration  $l$ . By Lemma 7,  $(\theta_j^{k_j^{l-1}})_{j \in B^{l-1}} \in \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$ . Consider the chosen bidder  $i$  whose price is incremented. The new price vector satisfies:

$$k_j^l := \begin{cases} k_j^{l-1} + 1 & \text{if } j = i \\ k_j^{l-1} & \text{otherwise.} \end{cases} \quad (109)$$

and

$$(\theta_j^{k_j^l})_{j \in B^{l-1}} \in \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q) \quad (110)$$

We now show that if bidder  $i$  quits, *i.e.* if his true type  $\theta_i$  is in the interval  $(\theta_i^{k_i^{l-1}}, \theta_i^{k_i^l}]$ , then he is required not to be in the allocation based on the information revealed thus far. We proceed by contradiction: If there exists  $\theta_{B^{l-1} \setminus i}^* > (\theta_j^{k_j^l})_{j \in B^{l-1} \setminus i}$  such that  $i \in f_y(\theta_i^{k_i^l}, \theta_{B^{l-1} \setminus i}^*, \theta_{N \setminus B^{l-1}}^Q)$ , then  $(\theta_j^{k_j^l})_{j \in B^{l-1}} \notin \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$ . (Thus contradicting Equation 110.)

Pick  $\epsilon$  such that  $0 < \epsilon < \min\{\theta_i^{k_i^l} - \theta_i^{k_i^{l-1}}, \min_{j \in B^{l-1} \setminus i}\{\theta_j^* - \theta_j^{k_j^l}\}\}$ . For all  $\theta_{B^{l-1}}^{**}$  such that  $\|\theta_{B^{l-1}}^{**} - (\theta_j^{k_j^l})_{j \in B^{l-1}}\|_2 < \epsilon$ , it follows (by the finite partition and by monotonicity) that  $i \in f_y(\theta_i^{**}, \theta_{B^{l-1} \setminus i}^*, \theta_{N \setminus B^{l-1}}^Q)$ , and moreover  $\theta_{B^{l-1} \setminus i}^{**} < \theta_{B^{l-1} \setminus i}^*$ . Thus,  $\theta_{B^{l-1}}^{**} \notin \{\theta_{B^{l-1}} : \forall \theta'_{B^{l-1} \setminus i} \geq \theta_{B^{l-1} \setminus i} : i \notin f_y(\theta_i, \theta'_{B^{l-1} \setminus i}, \theta_{N \setminus B^{l-1}}^Q)\}$ . Since this is true for all  $\theta_{B^{l-1}}^{**}$  in an  $\epsilon$ -ball around  $(\theta_j^{k_j^l})_{j \in B^{l-1}}$ , it follows that  $(\theta_j^{k_j^l})_{j \in B^{l-1}} \notin \text{closure}(\{\theta_{B^{l-1}} : \forall \theta'_{B^{l-1} \setminus i} \geq \theta_{B^{l-1} \setminus i} : i \notin f_y(\theta_i, \theta'_{B^{l-1} \setminus i}, \theta_{N \setminus B^{l-1}}^Q)\})$ . Thus  $(\theta_j^{k_j^l})_{j \in B^{l-1}} \notin \tilde{\Theta}_{B^{l-1}}(\theta_{N \setminus B^{l-1}}^Q)$ . We have the desired contradiction.

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  $\theta^*$ ,  $(B^l)^C \cap (y^l)^C \subseteq f_y(\theta^*)^C$ .

Gathering results: For all  $l$ ,  $y^l = (B^l)^C \cap (y^l)^C \subseteq f_y(\theta^*)$  and  $(B^l)^C \cap (y^l)^C \subseteq f_y(\theta^*)^C$ . Thus, whenever  $B^l = \emptyset$ ,  $f_y(\theta^*) = y^l$ . 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 18 - 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  $B \subseteq N$  and  $\theta_{N \setminus B}$ . We now show that  $\tilde{\Theta}_B(\theta_{N \setminus B})$  is connected. Let  $p : [0, 1] \rightarrow \Theta_B$  be the price path under  $G$  faced by agents in  $B$ , when the type profile for the agents in  $B$  is

$\sup\{\tilde{\Theta}_B(\theta_{N\setminus B})\}$  and the type profile for the agents in  $N \setminus B$  is  $\theta_{N\setminus B}$ . Let  $z$  be the terminal history that results from that type profile, and let  $l$  be the number of elements of that sequence. Let  $h_1, h_2, \dots, h_l$  be the subhistories of  $z$ .

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  $B$  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  $r$ ,  $p(r) \in \tilde{\Theta}_B(\theta_{N\setminus B})$ . Thus,  $p$  is a path from  $\tilde{\Theta}_B(\theta_{N\setminus B})$  to  $\sup\{\tilde{\Theta}_B(\theta_{N\setminus A})\}$ .

By Lemma 6,  $\tilde{\Theta}_B(\theta_{N\setminus B})$  is a join semi-lattice. We can generate a path  $p'$  from any  $\theta'_B \in \tilde{\Theta}_B(\theta_{N\setminus B})$  to  $\sup\{\tilde{\Theta}_B(\theta_{N\setminus B})\}$ , by defining  $p'(r) \equiv \theta'_B \vee p(r)$ . Thus,  $\tilde{\Theta}_B(\theta_{N\setminus B})$  is path-connected, which implies that it is connected.

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

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

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

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

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

$$p(\frac{t}{l}) \notin \text{closure}(\{\theta_B : \forall \theta'_{B \setminus i} \geq \theta_{B \setminus i} : i \notin f_y(\theta_i, \theta'_{B \setminus i}, \theta_{N\setminus B})\}) \quad (111)$$

The complement of the set on the RHS of Equation 111 is open. Thus, for some  $\epsilon > 0$ , an open  $\epsilon$ -ball around  $p(\frac{t}{l})$  is a subset of the complement of the RHS.

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

This completes the proof of the “only-if” direction. □

## B Alternative Empirical Specifications

Here we report alternative empirical specifications for the experiment. Table 5 and 6 are identical to Table 2 except that they compute  $p$ -values and standard errors using alternative methods.

A natural measure of errors would be to take the sum, for  $k = 1, 2, 3, 4$ , of the absolute difference between the  $k$ th highest bid and the  $k$ th highest value. However we do not observe the highest bid under **AC**, and we often do not observe the highest bid under **AC+X**. We could instead take the

Table 5: mean(abs(2nd bid - 2nd value),  $p$ -values calculated using Wilcoxon rank-sum test

Format	Rounds	SP	OSP	$p$ -value
Auction	1-5	8.04	3.19	< .001
	6-10	4.99	1.77	.005
+X Auction	1-5	3.99	1.83	.004
	6-10	3.69	1.29	< .001

This is the same as Table 2, except that the  $p$ -values are calculated using the Wilcoxon rank-sum test.

Table 6: mean(abs(2nd bid - 2nd value),  $p$ -values and standard errors calculated using clustered regression (clustered by groups)

Format	Rounds	SP	OSP	$p$ -value
Auction	1-5	8.04 (1.26)	3.19 (1.06)	.006
	6-10	4.99 (1.20)	1.77 (0.34)	.014
+X Auction	1-5	3.99 (0.61)	1.83 (0.41)	.005
	6-10	3.69 (0.88)	1.29 (0.33)	.015

This is the same as Table 2, except that the  $p$ -values and standard errors are calculated by running a single regression (with appropriate indicator variables) and clustering by groups.

sum for  $k = 2, 3, 4$  of the absolute difference between the  $k$ th highest bid and the  $k$ th highest value, averaged as before in five-round blocks. Table 7 reports the results.

Table 7: mean(sum(abs( $k$ th bid -  $k$ th value))), for  $k = 2, 3, 4$

Format	Rounds	SP	OSP	$p$ -value
Auction	1-5	32.63	9.89	< .001
		(4.64)	(1.89)	
	6-10	16.28	5.53	.001
		(2.73)	(0.91)	
+X Auction	1-5	17.04	6.18	.011
		(3.70)	(1.06)	
	6-10	14.21	4.74	.022
		(3.70)	(0.75)	

For each auction, we sum the absolute differences between the  $k$ th bid and the  $k$ th value, for  $k = 2, 3, 4$ . We then take the mean of this over each 5-round block. We then compute standard errors counting each group’s 5-round mean as a single observation. (18 observations per cell.)  $p$ -values are computed using a two-sample  $t$ -test, allowing for unequal variances.

Another measure of errors would be to take the sum of the absolute difference between each bidder’s bid and *that bidder’s* value, dropping all highest bidders for symmetry. Table 8 reports the results.

Table 9 reports the results of Table 3, except that the  $p$ -values are calculated using the Wilcoxon rank-sum test.

29.0% of rank-order lists are incorrect under **SP-RSD**. 2.6% of choices are incorrect under **OSP-RSD**. However, this is not a fair comparison; rank-order lists mechanically allow us to spot more errors than single choices. To compare like with like, we compute the proportion of incorrect choices we would have observed, if subjects played **OSP-RSD** as though they were implementing the submitted rank-order lists for **SP-RSD**. This is a cautious measure; it counts errors under **SP-RSD** only if they would have altered the outcome under **OSP-RSD**. Table 10 reports the results.

Finally, in Table 11, we compute the mean difference between the second-highest bid and the second-highest value, by auction format and by five-round blocks. This summarizes the average *direction* of deviations (positive or negative) from equilibrium play. I had no prior hypothesis about these outcome variables, but report this analysis for completeness.

## C Quantal Response Equilibrium

Quantal response equilibrium (QRE) is defined for normal form games (McKelvey and Palfrey, 1995). Agent quantal response equilibrium (AQRE) adapts QRE to extensive form games (McKelvey and Palfrey, 1998). In AQRE, each agent plays a (perturbed) best response at each information set, given correct beliefs about other agents’ play and his own mistakes at future information sets. In **AC** and **AC+X**, a rational agent is indifferent between quitting and continuing when the price is equal to his value, since if he continues presently, then he will quit at the next bid increment. Consequently, an AQRE agent *strictly* prefers to quit when the price is equal to his value, since he correctly foresees that if he continues, he will sometimes make a mistake at the next bid increment, and the increment after that, resulting in a negative expected value. By continuity, an AQRE agent strictly prefers to quit when the price is in some interval below his value. Consequently, AQRE

Table 8: mean(sum(abs( $i$ 's bid -  $i$ 's value))), dropping highest bidders

Format	Rounds	SP	OSP	$p$ -value
Auction	1-5	35.13 (5.20)	10.18 (1.88)	< .001
	6-10	15.46 (2.85)	4.89 (0.72)	.002
+X Auction	1-5	17.88 (4.10)	5.58 (1.01)	.009
	6-10	14.20 (3.72)	4.64 (0.83)	.022

For each auction, we sum the absolute differences between each bidder's bid and their value, dropping the highest bidder. We then take the mean of this over each 5-round block. We then compute standard errors counting each group's 5-round mean as a single observation. (18 observations per cell.)  $p$ -values are computed using a two-sample  $t$ -test, allowing for unequal variances.

Table 9: Proportion of serial dictatorships not ending in dominant strategy outcome,  $p$ -values calculated using Wilcoxon rank-sum test

	SP	OSP	$p$ -value
Rounds 1-5	43.3%	7.8%	.0001
Rounds 6-10	28.9%	6.7%	.0010

This is the same as Table 3, except that the  $p$ -values are calculated using the Wilcoxon rank-sum test.

Table 10: Proportion of incorrect choices under serial dictatorship: SP (imputed) vs OSP (actual)

	SP	OSP	$p$ -value
Rounds 1-5	17.8% (3.5%)	2.6% (1.1%)	< .001
Rounds 6-10	10.7% (2.0%)	2.6% (1.3%)	.002
$p$ -value	.078	1.000	

For each group in **SP-RSD**, for each period, we simulate the three choices that we would have observed under **OSP-RSD**. For each group, for each 5-round block, we record the proportion of choices that are incorrect. We then compute standard errors counting each group-block pair as a single observation. (18 observations per cell.) When comparing SP to OSP, we compute  $p$ -values using a two-sample  $t$ -test. When comparing early to late rounds of the same game, we compute  $p$ -values using a paired  $t$ -test. In the sample for **OSP-RSD**, there are 7 incorrect choices in the first five rounds and 7 incorrect choices in the last five rounds.

predicts systematic under-bidding in **AC** and **AC+X**.

As Figure 2 and Figure 3 illustrate, there is no evidence of systematic under-bidding in **AC** and **AC+X**. Table 11 shows this rigorously.

## D Experiment instructions



## WELCOME

This is a study about decision-making. Money earned will be paid to you in cash at the end of the experiment. This study is about 90 minutes long.

We will pay you \$5 for showing up, and \$15 for completing the experiment. Additionally, you will be paid in cash your earnings from the experiment. If you make choices in this experiment that lose money, we will deduct this from your total payment. However, your total payment (including your show-up payment and completion payment) will always be at least \$20.

You have been randomly assigned into groups of 4. This experiment involves 3 games played for real money. You will play each game 10 times with the other people in your group.

We will give you instructions about each game just before you begin to play it. Your choices in one game will not affect what happens in other games.

There is *no deception* in this experiment. Every game will be exactly as specified in the instructions. Anything else would violate the IRB protocol under which we run this study. (IRB Protocol 34876)

Please do not use electronic devices or talk with other volunteers during this study. If we do find you using electronic devices or talking with other volunteers, the rules of the study require us to deduct \$20 from your earnings.

If you have questions at any point, please raise your hand and we will answer your questions privately.

## GAME 1

In this game, you will bid in an auction for a money prize. The prize may have a different dollar value for each person in your group. You will play this game for 10 rounds. All dollar amounts in this game are in 25 cent increments.

At the start of each round, we display your value for this round's prize. If you win the prize, you will earn the value of the prize, minus any payments from the auction.

Your **value for the prize** will be calculated as follows:

1. For each group we will draw a **common value**, which will be between \$10.00 and \$100.00. Every number between \$10.00 and \$100.00 is equally likely to be drawn.
2. For each person, we will also draw a **private adjustment**, which will be between \$0.00 and \$20.00. Every number between \$0.00 and \$20.00 is equally likely to be drawn.

In each round, your **value for the prize** is equal to the **common value** plus your **private adjustment**. At the start of each round, you will learn your total value for the prize, but not the common value or the private adjustment.

This means that each person in your group may have a different value for the prize. However, when you have a high value, it is more likely that other people in your group have a high value.

The auction proceeds as follows: First, you will learn your value for the prize. Then you can choose a bid in the auction. Each person in your group will submit their bids privately and at the same time. You do this by typing your bid into a text box and clicking 'confirm bid'. You will have 90 seconds to make your decision, and can revise your bid as many times as you like. At the end of 90 seconds, your final bid will be the one that counts.

Your value for the prize	48.00
Current bid	0.00
Enter bid here:	<input type="text"/>
<input type="button" value="Confirm Bid"/>	

All bids must be between \$0.00 and \$150.00, and in 25 cent increments.

The highest bidder will win the prize, and make a payment equal to **the second-highest bid**. This means that we will **add** to her earnings her **value for the prize**, and **subtract** from her earnings **the second-highest bid**. All other bidders' earnings will not change.

At the end of each auction, we will show you the bids, ranked from highest to lowest, and the winning bidder's profits. If there is a tie for the highest bidder, no bidder will win the object.

## GAME 2

In this game, you will bid in an auction for a money prize. You will play this game for 10 rounds.

Your **value for the prize** will be generated as before.

However, each round, we will also draw a new number, **X**, for each group.

The rules of the auction are different, as follows:

All bidders will submit their bids privately and at once. However, the highest bidder will win the prize if and only if their bid **exceeds the second-highest bid by more than X**.

If the highest bidder wins the prize, she will make a payment equal to **the second-highest bid plus X**. This means that we will **add** to her earnings her **value for the prize**, and **subtract** from her earnings **the second-highest bid plus X**. All other bidders' earnings will not change.

If the highest bid does not exceed the second-highest bid by **more than X**, then no bidder will win the prize. In that case, no bidder's earnings will change.

**X** will be **between \$0.00 and \$3.00**, with every 25 cent increment equally likely to be drawn. You will be told your value for the prize at the start of each round, but will not be told **X**. At the end of each round, we will tell you the value of **X**.

### GAME 3

You will play this game for 10 rounds. In each round of this game, there are four prizes, labeled A, B, C, and D. Prizes will be worth between \$0.00 and \$1.25. For each prize, its value will be **the same for all the players** in your group.

At the start of each round, you will learn the value of each prize. You will also learn your priority score, which is a random number between 1 and 10. Every whole number between 1 and 10 is equally likely to be chosen.

The game proceeds as follows: We will ask you to list the prizes, **in any order of your choice**. All players will submit their lists privately and at the same time.

After all the lists have been submitted, we will assign prizes using the following rule:

1. The player with the highest priority score will be assigned the top prize on his list.
2. The player with the second-highest priority score will be assigned the top prize on his list, among the prizes that remain.
3. The player with the third-highest priority score will be assigned the top prize on his list, among the prizes that remain.
4. The player with the lowest priority score will be assigned whatever prize remains.

If two players have the same priority score, we will break the tie randomly.

You will have 90 seconds to form your list. You do this by typing a number, from 1 to 4, next to each prize, and then clicking the button that says “Confirm Choices”. Each prize must be assigned a different number, from 1 (top) to 4 (bottom). Your choices will not count unless you click the button that says “Confirm Choices”.

Prize	Value (\$)	Choose (1-4)	Rank
A	0.75	<input type="text" value="1"/>	
B	0.50	<input type="text"/>	
C	1.25	<input type="text"/>	
D	0.00	<input type="text"/>	

Your priority score is 5.  
Rank the prizes in any order from 1 to 4.

[Confirm Choices](#)

If you do not produce a list by the end of 90 seconds, we will assign prizes as though you reported the list in order A-B-C-D.

At the end of each round, we will **add** to your earnings the value of the prize you were assigned.

## WELCOME

This is a study about decision-making. Money earned will be paid to you in cash at the end of the experiment. This study is about 90 minutes long.

We will pay you \$5 for showing up, and \$15 for completing the experiment. Additionally, you will be paid in cash your earnings from the experiment. If you make choices in this experiment that lose money, we will deduct this from your total payment. However, your total payment (including your show-up payment and completion payment) will always be at least \$20.

You have been randomly assigned into groups of 4. This experiment involves 3 games played for real money. You will play each game 10 times with the other people in your group.

We will give you instructions about each game just before you begin to play it. Your choices in one game will not affect what happens in other games.

There is *no deception* in this experiment. Every game will be exactly as specified in the instructions. Anything else would violate the IRB protocol under which we run this study. (IRB Protocol 34876)

Please do not use electronic devices or talk with other volunteers during this study. If we do find you using electronic devices or talking with other volunteers, the rules of the study require us to deduct \$20 from your earnings.

If you have questions at any point, please raise your hand and we will answer your questions privately.

## GAME 1

In this game, you will bid in an auction for a money prize. The prize may have a different dollar value for each person in your group. You will play this game for 10 rounds. All dollar amounts in this game are in 25 cent increments.

At the start of each round, we display your value for this round's prize. If you win the prize, you will earn the value of the prize, minus any payments from the auction.

Your **value for the prize** will be calculated as follows:

1. For each group we will draw a **common value**, which will be between \$10.00 and \$100.00. Every number between \$10.00 and \$100.00 is equally likely to be drawn.
2. For each person, we will also draw a **private adjustment**, which will be between \$0.00 and \$20.00. Every number between \$0.00 and \$20.00 is equally likely to be drawn.

In each round, your **value for the prize** is equal to the **common value** plus your **private adjustment**. At the start of each round, you will learn your total value for the prize, but not the common value or the private adjustment.

This means that each person in your group may have a different value for the prize. However, when you have a high value, it is more likely that other people in your group have a high value.

The auction proceeds as follows: First, you will learn your value for the prize. Then, the auction will start. We will display a price to everyone in your group, that starts low and counts upwards in 25 cent increments, up to a maximum of \$150.00. At any point, you can choose to leave the auction, by clicking the button that says "Stop Bidding".

ACTIVE BIDDERS	4
PRICE	35.25
Your value for the prize	48.00
Stopped bidding at	



**When there is only one bidder left in the auction**, that bidder will win the prize at **the current price**. This means that we will **add** to her earnings her **value for the prize**, and **subtract** from her earnings **the current price**. All other bidders' earnings will not change.

At the end of each auction, we will show you the prices where bidders stopped, and the winning bidder's profits. If there is a tie for the highest bidder, no bidder will win the object.

## GAME 2

In this game, you will bid in an auction for a money prize. You will play this game for 10 rounds.

Your **value for the prize** will be generated as before.

However, each round, we will also draw a new number, **X**, for each group.

The rules of the auction are different, as follows:

The price will count up from a low value, and you can choose to leave the auction at any point, by clicking the button that says "Stop Bidding". When there is only one bidder left in the auction, the price will **continue to rise for another X dollars**, and then freeze.

If the last bidder **stays in the auction until the price freezes**, then she will win the prize at **the final price**. This means that we will **add** to her earnings her **value for the prize**, and **subtract** from her earnings **the final price**. All other bidders' earnings will not change.

If the last bidder stops bidding before the price freezes, then no bidder will win the prize. In that case, no bidder's earnings will change.

**X** will be **between \$0.00 and \$3.00**, with every 25 cent increment equally likely to be drawn. You will be told your value for the prize at the start of each round, but will not be told **X**. At the end of each round, we will tell you the value of **X**.

### GAME 3

You will play this game for 10 rounds. In each round of this game, there are four prizes, labeled A, B, C, and D. Prizes will be worth between \$0.00 and \$1.25. For each prize, its value will be **the same for all the players** in your group.

At the start of each round, you will learn the value of each prize. You will also learn your priority score, which is a random number between 1 and 10. Every whole number between 1 and 10 is equally likely to be chosen.

The game proceeds as follows:

1. The player with the highest priority score will pick one prize.
2. The player with the second-highest priority score will pick one of the prizes that remains.
3. The player with the third-highest priority score will pick one of the prizes that remains.
4. The player with the lowest priority score will be assigned whatever prize remains.

If two players have the same priority score, we will break the tie randomly.

When it is your turn to pick, you will have 30 seconds to make your choice. You do this by selecting a prize and then clicking the button that says “Confirm Choice”. Your choices will not count unless you click the button that says “Confirm Choice”.

Prize	Value (\$)
A	0.75
B	0.50
C	1.25
D	0.00

Pick one of the prizes

A

B

C

D

**Confirm Choice**

If you do not make a choice by the end of 30 seconds, we will assign prizes as though you picked whichever prize is earliest in the alphabet.

At the end of each round, we will **add** to your earnings the value of the prize you picked.

Table 11: mean(2nd bid - 2nd value)

Format	Rounds	SP	OSP
Auction	1-5	-1.59 (2.07)	0.58 (0.77)
	6-10	4.22*** (1.28)	0.78* (0.40)
+X Auction	1-5	1.62** (0.70)	0.15 (0.47)
	6-10	2.97*** (0.97)	0.02 (0.31)

For each group, we take the mean difference between the second-highest bid and the second-highest value over each 5-round block. We then compute standard errors counting each group's 5-round mean as a single observation. (18 observations per cell, standard errors in parentheses.) We then run a two-sided  $t$ -test for the null that the value in each cell has zero mean. \* denotes  $p < .10$ , \*\* denotes  $p < .05$ , and \*\*\* denotes  $p < .01$ .