Obvious Dominance and Random Priority

Marek Pycia and Peter Troyan*

January 2018

Abstract

We construct the full class of obviously strategy-proof mechanisms in environments without transfers as the class of clinch-or-pass games we call millipede games. Some millipede games are indeed simple and widely used in practice, while others may be complex, requiring agents to perform lengthy backward induction, and are rarely observed. We introduce a natural strengthening of obvious strategy-proofness called strong obvious strategy-proofness, which eliminates some of the more complex millipede games. We use our definition to characterize the well-known Random Priority mechanism as the unique mechanism that is efficient, fair, and simple to play, thereby explaining its popularity in practical applications.

1 Introduction

Consider a group of agents who must come together to make a choice from some set of potential outcomes that will affect each of them. This can be modeled as having the agents play a "game", taking turns choosing from sets of actions (possibly simultaneously), with the final outcome determined by the decisions made by all of the agents each time they were called to play. To ensure that the ultimate decision taken satisfies desirable normative properties (e.g., efficiency), the incentives given to the agents are crucial. The standard route taken in mechanism design to ensure good incentives is to appeal to the revelation principle and look for direct mechanisms that are strategy-proof, i.e., mechanisms where agents are

*Pycia: UCLA and U. Zurich; Troyan: University of Virginia. First presentation: April 2016. First posted draft: June 2016. For their comments, we would like to thank Itai Ashlagi, Eduardo Azevedo, Ben Golub, Yannai Gonczarowski, Ed Green, Fuhito Kojima, Shengwu Li, Giorgio Martini, Stephen Morris, Erling Skancke, Utku Unver, the Eco 514 students at Princeton, and the audiences at the NBER Market Design workshop, NEGT'16, NC State, ITAM, NSF/CEME Decentralization, and the Econometric Society Meetings. Simon Lazarus provided excellent research assistance. Pycia would like to acknowledge the financial support of the William S. Dietrich II Economic Theory Center at Princeton.

simply asked to report their private information, and it is always in their interest to do so truthfully, no matter what the other agents do. However, this is useful only to the extent the participants *understand* that a given mechanism is strategy-proof, and indeed, there is evidence many real-world agents do not tell the truth, even in strategy-proof mechanisms. In other words, strategy-proof mechanisms, while theoretically appealing, may not actually be easy for participants to play in practice. What mechanisms, then, are actually "simple to play"? And further, what are the trade-offs among simplicity and other normatively desirable properties such as efficiency and fairness? This paper provides answers to these questions.

We begin by constructing the full class of *obviously strategy-proof* (Li, 2017), or OSP, mechanisms in general social choice environments without transfers. Social choice problems without transfers are ubiquitous in the real-world, and examples include refugee resettlement, school choice, organ exchange, public housing allocation, course allocation, and voting, among others.[1] We call the class of OSP games in these environments *millipede games* (for reasons that will become clear shortly). While some millipede games, such as sequential dictatorships, are frequently encountered and are indeed simple to play, others are rarely observed in market-design practice, and their strategy-proofness is not necessarily immediately clear. In particular, some millipede games may still require agents to look far into the future and to perform potentially complicated backward induction reasoning. Thus, to further delineate the class of mechanism that are simple to play, we introduce a new criterion called *strong obvious strategy-proofness (SOSP)*. We construct the full class of strongly obviously strategy-proof games and show that strong obvious strategy-proofness selects the subset of millipede games that are observed in practice while eliminating the more complex millipede games that are rarely (if ever) used. Combining SOSP with standard efficiency and fairness axioms narrows the space of mechanisms down even further to a unique choice: Random Priority, a mechanism that is commonly observed in practical applications.

An imperfect-information extensive-form game is obviously strategy-proof if, whenever an agent is called to play, there is an action such that even the worst possible final outcome from following the given action is at least as good as the best possible outcome from taking any other action at the node in question, where the best and worst case are determined by assuming that the agent plays optimally and considering all possible strategies that could be played by the agent's opponents in the future. Li (2017) provides both a theoretical behavioral foundation for OSP games being "simple to play" and experimental evidence that, in certain settings, OSP mechanisms do indeed lead to higher rates of truth-telling in

[1] See e.g. Roth (2015) and Jones and Teytelboym (2016) for resettlement, Abdulkadiroğlu and Sönmez (2003) for school choice, Roth, Sönmez, and Ünver (2004) for transplants, Sönmez and Ünver (2010) and Budish and Cantillon (2012) for course allocation, and Arrow (1963) for voting and social choice.

practice than their counterparts that are strategy-proof, but not obviously so.[2]

Obvious strategy-proofness will clearly restrict the set of potential games from which a designer may choose, which raises the question of exactly which games are obviously strategy-proof. Our first main result, Theorem 1, constructs the full class of OSP mechanisms as the class of millipede games. In a millipede game, Nature moves first and chooses a deterministic subgame, after which agents engage in a game of passing and clinching that resembles the well-studied centipede game (Rosenthal, 1981). To describe this deterministic subgame in this introduction, for expositional ease, we focus on allocation problems with agents who demand at most one unit; the argument for more general social choice environments is similar. An agent is presented with some subset of objects that she can 'clinch', or, take immediately and leave the game; she also may be given the opportunity to 'pass', and remain in the game.[3] If this agent passes, another agent is presented with an analogous choice. Agents keep passing among themselves until one of them clinches some object. When an agent clinches an object, this is her last move and she will be allocated this object at the end of the game.

The class of millipede games includes many games that are commonly seen in practice, such as Random Priority and Serial Dictatorships, which are millipede games in which the agent who moves can always clinch any object that is still available. However, our characterization shows that these are not the only mechanisms that are obviously strategy-proof. Some millipede games require substantial foresight on the part of the agents, similar to the foresight required in centipede games. For instance, it is possible to construct a millipede game such that a player is offered the possibility of clinching his second-choice object, but not his top choice object, even though it is still available. If the agent passes, he might not be given the opportunity to clinch any of his top fifty objects in the next one hundred moves. The obviously dominant strategy requires the agent to pass, even though he may never be offered his top choice for some plays of the other agents. (We provide an example of such a game in Figure 2.) Recognizing that passing is obviously dominant requires the agent to perform lengthy backward induction.

As another illustration of the backward-induction requirements inherent in some obviously dominant strategies consider chess. If White can always force a win, then any winning strategy of White is obviously dominant, yet the strategic choices in chess are far from obvi-

[2]He also shows that OSP games are those that can be implemented under bilateral commitment between the designer and the agents.

[3]In general, we may allow this agent to clinch the same object in several ways; while the choice between them has no impact on the agent's outcome, it might affect the allocation of others. Additionally, there need not be a passing action that allows the agent to stay in the game, but if there is, the main restriction implied by OSP is that there can be at most one such action.

ous. The reason is that obvious dominance allows for agents who are unable to contingently reason about their opponents' actions, but presumes that they do have perfect foresight about their own future selves. Thus, if White can force a win, then, at each node, he is able to determine the obviously dominant move by reasoning backwards from all future nodes at which he is called to play. Indeed, to the best of our knowledge even the question whether White actually has a winning strategy remains open.[4]

These examples suggest that some obviously dominant strategies may be hard to identify, which poses the question of what properties of an extensive-form game guarantee that the game is simple to play. To address this question, we introduce a refinement of obvious strategy-proofness, which we call *strong obvious strategy-proofness.* An extensive-form game is strongly obviously strategy-proof if, whenever an agent is called to play, there is an action such that even the worst possible final outcome from that action is at least as good as the best possible outcome from any other action, where what is possible may depend on all future actions, including actions by the agent's future-self. Thus, strongly obviously dominant strategies are those that are weakly better than all alternative strategies even if the agent is concerned that she might tremble in the future or has time-inconsistent preferences. Further, all SOSP games can be correctly played by agents who may not be able to look far into the future and perform lengthy backwards induction.

We show this by proving that strong obvious strategy-proofness eliminates the complex members of the more general class of millipede games, and that strongly obviously strategy-proof games take the form of curated dictatorships. In curated dictatorships, all agents move only once and (with the possible exception of the penultimate mover) their outcomes do not depend on the choices of agents who move after them. The optimal play thus relies on the agents's ability to see at most one step ahead in the game.[5]

While most of our focus is on incentives and simplicity, there are other important criteria when designing a mechanism, such as efficiency and fairness. Our final result shows when these desiderata are added as necessary requirements, only one mechanism survives: Random Priority. Random Priority is extensively used in a wide variety of practical allocation problems. School choice, worker assignment, course allocation, and the allocation of public housing are just a few of many examples, both formal and informal. Random Priority is well-known to have good efficiency, fairness, and incentive properties.[6] However, it has until now remained unknown whether there are other such mechanisms, and if so, what explains

[4] We would like to thank Eduardo Azevedo and Ben Golub for raising this point.

[5] In particular, strongly obviously strategy-proof and Pareto efficient games take the simple form of almost-sequential dictatorships, studied earlier in a different context by Pycia and Ünver (2016).

[6] For discussion of efficiency and fairness see, e.g., Abdulkadiroğlu and Sönmez (1998), Bogomolnaia and Moulin (2001), and Che and Kojima (2010).

the relative popularity of RP over these alternatives.[7] Theorem 5 provide answers to these questions by showing that there is no other such mechanism: a game is strongly obviously strategy-proof, Pareto efficient, and treats agents equally (a standard, and relatively weak, fairness axiom) if and only if it is Random Priority. This insight resolves positively the quest to establish Random Priority as the unique mechanism with good incentive, efficiency, and fairness properties, thereby explaining its popularity in practical market design settings.

Our results build on the key contributions of Li (2017), who formalized obvious strategy-proofness and established its desirability as an incentive property by both providing a theoretical foundation and experimental evidence that participants play their dominant strategy more often in obviously strategy-proof mechanisms (e.g., ascending auctions, dynamic Random Priority) than in mechanisms that are strategy-proof, but not obviously so (sealed-bid auctions, static Random Priority).[8] While Li looks at specific mechanisms, we characterize the entire class of obviously strategy-proof mechanisms, and provide an explanation for the popularity of Random Priority over all other mechanisms, results which have no counterpart in his work. Finally, our analysis of strong obvious strategy-proofness furthers our understanding of why some extensive forms of a mechanism are more often encountered in practice, despite being both equivalent according to the standard Myerson-Riley revelation principle and obviously strategy-proof.

Following up on Li's work, but preceding ours, Ashlagi and Gonczarowski (2016) show that stable mechanisms such as Deferred Acceptance are not obviously strategy-proof, except in very restrictive environments where Deferred Acceptance simplifies to an obviously strategy-proof game with a 'clinch or pass' structure similar to simple millipede games (though they do not describe it in these terms). Troyan (2016) studies obviously strategy-proof allocation via the popular Top Trading Cycles (TTC) mechanism, and provides a characterization of the priority structures under which TTC is OSP-implementable.[9]

[7] In single-unit demand allocation with at most three agents and three objects, Bogomolnaia and Moulin (2001) proved that Random Priority is the unique mechanism that is strategy-proof, efficient, and symmetric. In markets in which each object is represented by many copies, Liu and Pycia (2013) and Pycia (2011) proved that Random Priority is the asymptotically unique mechanism that is symmetric, asymptotically strategy-proof, and asymptotically ordinally efficient. While these earlier results looked at either very small or very large markets, ours is the first characterization that holds for any number of agents and objects.

[8] Li also shows that the classic top trading cycles (TTC) mechanism of Shapley and Scarf (1974), in which each agent starts by owning exactly one object, is not obviously strategy-proof. Troyan (2016) expands this by considering more general ownership structures, and shows that TTC is obviously strategy-proof if and only if at any time, at most two agents own all of the available objects. Also of note is Loertscher and Marx (2015) who study environments with transfers and construct a prior-free obviously strategy-proof mechanism that becomes asymptotically optimal as the number of buyers and sellers grows.

[9] Following on our work, Arribillaga et al. (2016) study obviously strategy-proof voting rules and Bade and Gonczarowski (2016) study obviously strategy-proof and efficient social choice rules in several environments.

More generally, this paper adds to our understanding of incentives, efficiency, and fairness in settings without transfers. In addition to Gibbard (1973, 1977) and Satterthwaite (1975), and the allocation papers mentioned above, the literature on mechanisms satisfying these key objectives includes Pápai (2000), Ehlers (2002) and Pycia and Ünver (2009) who characterized efficient and group strategy-proof mechanisms in settings with single-unit demand, and Pápai (2001) and Hatfield (2009) who provided such characterizations for settings with multi-unit demand.[10] Liu and Pycia (2013), Pycia (2011), Morrill (2014), and Hakimov and Kesten (2014) characterize mechanisms that satisfy incentive, efficiency, and fairness objectives.

2 Model

Let $\mathcal{N} = \{i_1, \ldots, i_N\}$ be a set of agents, and $\mathcal{X}$ a finite set of outcomes.[11] Each agent has a preference ranking over outcomes, where we write $x \succsim_i y$ to denote that x is weakly preferred to y. We allow for indifferences, and write $x \sim_i y$ if $x \succsim_i y$ and $y \succsim_i x$. The domain of preferences of agent $i \in \mathcal{N}$ is denoted $\mathcal{P}_i$, where each $\mathcal{P}_i$ determines a partition over the set of outcomes such that, for all $x, y \in \mathcal{X}$ that belong to the same element of the partition, we have $x \sim_i y$ for all $\succsim_i \in \mathcal{P}_i$. In particular, each preference relation of agent i may then be associated with the corresponding strict ranking of the elements of the partition. We will generally work with this strict ranking, denoted $\succ_i$, and will sometimes refer to $\succ_i$ as an agent's *type*.

The main assumption we make on the preference domains is that they are rich, in the following sense: given a partition, every strict ranking of the elements of the partition is in $\mathcal{P}_i$. This assumption is satisfied for a wide variety of preference structures, including, for example, the canonical voting environment, where every agent can strictly rank all alternatives, as well as allocation problems in which each agent cares only about his or her own assignment. In the former case, each agent partitions $\mathcal{X}$ into $|\mathcal{X}|$ singleton subsets and an agent's type $\succ_i$ is a strict preference relation over $\mathcal{X}$. In the latter case, agents are indifferent between how objects she does not receive are assigned to others, and so each element of agent i's partition

Mackenzie (2017) introduces the notion of a "round table mechanism" for OSP implementation and draws parallels with the standard Myerson-Riley revelation principle for direct mechanisms.

[10] Pycia and Ünver (2016) characterized individually strategy-proof and Arrovian efficient mechanisms. For an analysis of these issues under additional feasibility constraints, see also Dur and Ünver (2015).

[11] The assumption that $\mathcal{X}$ is finite simplifies the exposition and it is satisfied in such examples of our setting as voting and the no-transfer allocation environments listed in the introduction. This assumption can be relaxed. For instance, our analysis goes through with no substantive changes if we allow infinite $\mathcal{X}$ endowed with a topology such that agents' preferences are continuous in this topology and the relevant sets of outcomes are compact.

of $\mathcal{X}$ can be identified with her own allocation, and her type $\succ_i$ is then the (strict) rankings of these allocations. Where this assumption fails is in settings with transfers, in which each agent always prefers having more money to less.

When dealing with lotteries, we are agnostic as to how agents evaluate them, as long as the following property holds: an agent prefers lottery μ over ν if for any outcomes $x \in \operatorname{supp}(\mu)$ and $y \in \operatorname{supp}(\nu)$ this agent weakly prefers x over y; the preference between μ and ν is strict if, additionally, at least one of the preferences between $x \in \operatorname{supp}(\mu)$ and $y \in \operatorname{supp}(\nu)$ is strict. This mild assumption is satisfied for expected utility agents; it is also satisfied for agents who prefer μ to ν as soon as μ first-order stochastically dominates ν.

To determine the outcome that will be implemented, the planner designs a game Γ for the agents to play. Formally, we consider imperfect-information, extensive-form games with perfect recall, which are defined in the standard way: there is a finite collection of partially ordered *histories* (sequences of moves), $\mathcal{H}$. At every non-terminal history $h \in \mathcal{H}$, one agent is called to play and has a finite set of *actions* $A(h)$ from which to choose. We allow for chance moves by Nature, and at any history h at which Nature is called to play, we use $\Omega(h)$ to denote the probability distribution over Nature's possible actions $A(h)$. We use the notation $\omega := (\omega(h))_{\{h\in\mathcal{H}:\text{nature moves at } h\}}$ to denote one particular realization of Nature's moves throughout the game at every history at which Nature is called to play. Each terminal history is associated with an outcome in $\mathcal{X}$, and agents receive payoffs at each terminal history that are consistent with their preferences over outcomes $\succ_i$.

We use the notation $h' \subseteq h$ to denote that h' is a subhistory of h (equivalently, h is a continuation history of h'), and write $h \subset h'$ when $h \subseteq h'$ but $h \neq h'$. $\mathcal{H}_i(h)$ denotes the set of (strict) subhistories $h' \subset h$ at which agent i is called to move. When useful, we sometimes write $h' = (h, a)$ to denote the history h' that is reached by starting at history h and following the action $a \in A(h)$.

An *information set* $\mathcal{I}$ of agent i is a set of histories such that for any $h, h' \in \mathcal{I}$ and any subhistories $\tilde{h} \subseteq h$ and $\tilde{h}' \subseteq h'$ at which i moves at least one of the following two symmetric conditions obtains: either (i) there is a history $\tilde{h}^* \subseteq \tilde{h}$ such that $\tilde{h}^*$ and $\tilde{h}'$ are in the same information set, $A(\tilde{h}^*) = A(\tilde{h}')$, and i makes the same move at $\tilde{h}^*$ and $\tilde{h}'$, or (ii) there is a history $\tilde{h}^* \subseteq \tilde{h}'$ such that $\tilde{h}^*$ and $\tilde{h}$ are in the same information set, $A(\tilde{h}^*) = A(\tilde{h})$, and i makes the same move at $\tilde{h}^*$ and $\tilde{h}$. We denote by $\mathcal{I}(h)$ the information set containing history h.[12] These imperfect information games allow us to incorporate incomplete information in the standard way in which Nature moves first and determines agents' types.

[12]It will turn out to be without loss of generality to assume all information sets are singletons, and so we will be able to drop the $\mathcal{I}(h)$ notation and identify each information set with the unique sequence of actions (i.e., history) taken to reach it.

A *strategy* for a player i in game Γ is a function S_i that specifies an action at each one of her information sets.[13] When we want to refer to the strategies of different types $\succ_i$ of agent i, we write $S_i(\succ_i)$ for the strategy followed by agent i for whom Nature drew type $\succ_i$; in particular, $S_i(\succ_i)(\mathcal{I})$ denotes the action chosen by agent i with type $\succ_i$ at information set $\mathcal{I}$. We use $S_{\mathcal{N}}(\succ_{\mathcal{N}}) = (S_i(\succ_i))_{i\in\mathcal{N}}$ to denote the strategy profile for all of the agents when the type profile is $\succ_{\mathcal{N}}= (\succ_i)_{i\in\mathcal{N}}$. An *extensive-form mechanism*, or simply a *mechanism,* is an extensive-form game Γ together with a profile of strategies $S_{\mathcal{N}}$. Two extensive-form mechanisms $(\Gamma, S_{\mathcal{N}})$ and $(\Gamma', S'_{\mathcal{N}})$ are *equivalent* if for every profile of types $\succ_{\mathcal{N}}= (\succ_i)_{i\in\mathcal{N}}$, the resulting distribution over outcomes when agents follow $S_{\mathcal{N}}(\succ_{\mathcal{N}})$ in Γ is the same as when agents follow $S'_{\mathcal{N}}(\succ_{\mathcal{N}})$ in Γ'.

Remark 1. In the sequel we establish several equivalences. Each of them is an equivalence of two mechanism which additionally satisfy additional criteria such as obvious strategy-proofness or strong obvious strategy-proofness. We can thus formally strengthen these results to $\mathcal{C}$-equivalence defined as follows. A *solution concept* $\mathcal{C}(\cdot)$ maps any game Γ into a subset of strategy profiles $\mathcal{C}(\Gamma)$. The interpretation is that the strategy profiles in $\mathcal{C}(\Gamma)$ are those profiles that satisfy the solution concept $\mathcal{C}$; for example, if we were concerned with strategy-proof implementation ($\mathcal{C}$ = SP), then SP(Γ) would be all strategy profiles $S_{\mathcal{N}}$ such that $S_i(\succ_i)$ is a weakly dominant strategy for all i and all types $\succ_i$ in game Γ. Two extensive-form mechanisms $(\Gamma, S_{\mathcal{N}})$ and $(\Gamma', S'_{\mathcal{N}})$ are $\mathcal{C}-$*equivalent* if the following two conditions are satisfied: (i) for every profile of types $\succ_{\mathcal{N}}= (\succ_i)_{i\in\mathcal{N}}$, the resulting distribution over outcomes when agents play $S_{\mathcal{N}}(\succ_{\mathcal{N}})$ in Γ is the same as when agents play $S'_{\mathcal{N}}(\succ_{\mathcal{N}})$ in Γ' and (ii) $S_{\mathcal{N}}(\succ_{\mathcal{N}}) \in \mathcal{C}(\Gamma)$ and $S'_{\mathcal{N}}(\succ_{\mathcal{N}}) \in \mathcal{C}(\Gamma')$. In other words, two mechanisms are $\mathcal{C}$-equivalent if they are equivalent and the strategies satisfy the solution concept $\mathcal{C}$ in the respective games.

3 Millipede Games

In this section, we begin our analysis of which games are "simple to play" by characterizing the entire class of games that are obviously strategy-proof. Following Li (2017), given a game Γ, a strategy S_i *obviously dominates* another strategy S'_i for player i if, starting from any earliest information set $\mathcal{I}$ at which these two strategies diverge,[14] the worst possible payoff to the agent from playing S_i is at least as good as the best possible payoff from S'_i, where the best/worst case outcomes are determined over all possible (S_{-i}, ω). A profile of

[13] We consider pure strategies, but the analysis can be easily extended to mixed strategies.

[14] That is, information set $\mathcal{I}$ is on the path of play under both S_i and S'_i and both strategies choose the same action at all earlier information sets but choose a different action at $\mathcal{I}$. Li (2017) refers to such an information set as an *earliest point of departure.* Note that for two strategies, there will in general be multiple earliest points of departure.

strategies $S_{\mathcal{N}}(\cdot) = (S_i(\cdot))_{i \in \mathcal{N}}$ is *obviously dominant* if for every player i and every type $\succ_i$, the strategy $S_i(\succ_i)$ obviously dominates any other strategy. When there exists a profile of strategies $S_{\mathcal{N}}(\cdot)$ that is obviously dominant, we say Γ is *obviously strategy-proof (OSP)*.

Obvious strategy-proofness is one way of capturing what it means for a mechanism to be simple to play. Li (2017) provides a theoretical foundation for this intuition by showing that obviously dominant strategies are those that can be recognized as dominant by cognitively limited agents who are unable to engage in contingent reasoning.[15] Further, he provides empirical evidence that, for certain mechanisms, participants are more likely to tell the truth under extensive-form implementations (which are OSP) than under the corresponding normal form (which are strategy-proof, but not obviously so).[16]

If obvious strategy-proofness as a solution concept does indeed capture what games are simple to play, then an important question is to determine precisely which games satisfy this criterion. Our first main result is to characterize the entire class of obviously strategy-proof mechanisms as a class of games that we call "millipede games". Intuitively, a millipede game is a take-or-pass game similar to a centipede game, but with more players and more actions (i.e., "legs") at each node. Figure 1 shows the extensive form of a millipede game for the special case of object allocation with single-unit demand, where the agents are labeled $i, j, k, \ldots$ and the objects are labeled $w, x, y, \ldots$. At the start of the game, the first mover, agent i has three options: he can take x, take y, or pass to agent j.[17] If he takes an object, he leaves the game and it continues with a new agent. If he passes, then agent j can take x, take z, or pass back to i. If he passes back to i, then i's possible choices increase from his previous move (he can now take z). The game continues in this manner until all objects have been allocated.

While Figure 1 considers an object allocation environment, millipede games can be defined more generally on any preference structure that satisfies our assumptions of Section 2. Recall that each agent's preference domain $\mathcal{P}_i$ partitions the outcome space $\mathcal{X}$ into indifference classes. We use the term *payoff* to refer to the indifference class associated with a particular element of the partition. We say that a payoff x is *possible* for agent i at history h if there is a strategy profile of all the agents (including choices made by Nature) such that

[15] See also Zhang and Levin (2017), who provide further decision-theoretic foundations for obvious dominance.

[16] For example, people are more likely to tell the truth in the extensive-form of the Random Priority mechanism than in the corresponding normal form (where they are asked to submit an entire list of preferences to 'the mechanism' ex-ante). From the standpoint of the classical revelation principle, these two mechanisms should be strategically equivalent.

[17] In the general definition of a millipede game, it will be possible that none of the actions are passing actions and so all actions are taking actions. The key restriction imposed by obvious strategy-proofness is that if there is a passing action, there can only be one. A formal definition is given below.

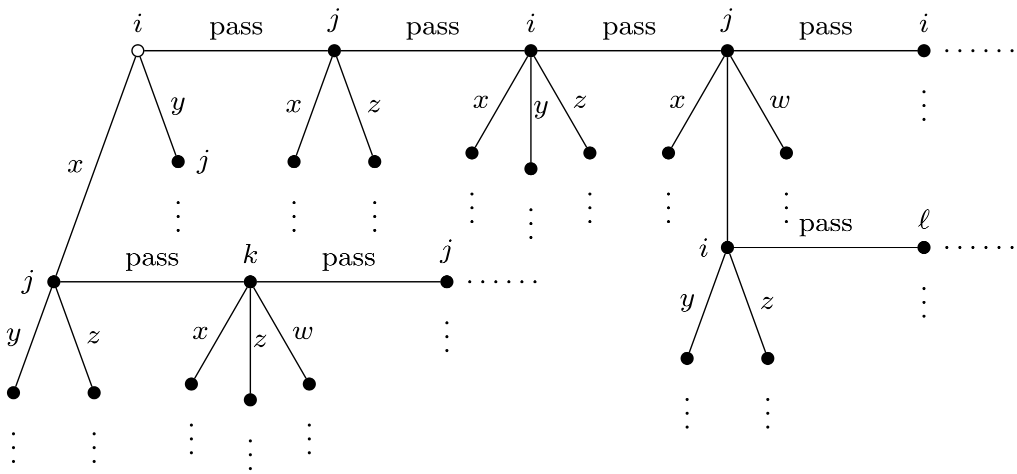

Figure 1: An example of a millipede game in the context of object allocation.

h is on the path of the game and, under this strategy profile, agent i obtains payoff x (i.e., the outcome that obtains under the given strategy profile is in the indifference class that gives her payoff x). For any history h, $P_i(h)$ denotes the set of payoffs that are possible for i at h. We say agent i has *clinched* payoff x at history h if agent i receives payoff x at all terminal histories $\bar{h} \supseteq h$. At a history h, if, by taking some action $a \in A(h)$ an agent never moves again in the game and receives the same payoff for every terminal $\bar{h} \supseteq (h, a)$, we say that a is a *clinching* action. Clinching actions are generalizations of the "taking actions" of Figure 1 to environments where the outcomes/payoff structure may be different from object allocation (where more generally, what i is clinching is a particular indifference class for herself). We denote the set of payoffs that i can clinch at history h by $C_i(h)$.[18] If an action $a \in A(h)$ is not a clinching action, then it is called a *passing* action.

A **millipede game** is a finite extensive-form game of perfect information that satisfies the following properties: Nature either moves once, at the empty history $\emptyset$, or Nature has no moves. At any history h at which an agent, say i, moves, all but at most one action are clinching actions; the remaining action (if there is one) is a passing action.[19] And finally, for all i, all histories h at which i moves and all terminal histories, and all payoffs x, at least one of the following holds:[20] (a) $x \in P_i(h)$; or (b) $x \notin P_i(\tilde{h})$ for some $\tilde{h} \in \mathcal{H}_i(h)$; or (c) $x \in \cup_{\tilde{h} \in \mathcal{H}_i(h)} C_i(\tilde{h})$; or (d) $\cup_{\tilde{h} \in \mathcal{H}_i(h)} C_i(\tilde{h}) \subseteq C_i(h)$. In words, these last conditions require

[18] That is, $x \in C_i(h)$ if there is some action $a \in A(h)$ such that i receives payoff x for all terminal $\bar{h} \supseteq (h, a)$. At a terminal history $\bar{h}$, no agent is called to move and there are no actions; however, for the purposes of constructing millipede games below, it will be useful to define $C_i(\bar{h}) = \{x\}$ for all i, where x is the payoff associated with the unique outcome that obtains at terminal history $\bar{h}$.

[19] There may be several clinching actions associated with the same final payoff.

[20] Recall that $\mathcal{H}_i(h) = \{h' \subsetneq h : i \text{ moves at } h'\}$.

that, for every payoff x that was at some point possible for i, either: (a) x remains possible at h, or (b) it was not possible already at an earlier history, or (c) it was clinchable at an earlier history, or (d) each payoff clinchable at an earlier history is clinchable at h.

Conditions (a)-(d) ensure that, if an agent's top possible outcome is not clinchable at some history h, then the payoff she ultimately receives must be at least as good as any payoff she could have clinched at h, which is necessary for passing to be obviously dominant. To see this, consider an agent whose top payoff is x, and a history h such that (a), (b) and (c) fail for x. This means at all prior moves, x was possible (by "not (b)"), but not clinchable (by "not (c)"), and so obvious dominance requires i to pass. However, x has disappeared as a possibility at h (by "not (a)"), and so to ensure that i does not regret her choice to pass at an earlier history, at h, we must offer her the opportunity to clinch anything she could have clinched previously, which is condition (d).

Notice that millipede games have a recursive structure: the continuation game that follows any action is also a millipede game. A simple example of a millipede game is a deterministic serial dictatorship in which no agent ever passes and there is only one active agent at each node.[21] A more complex example is given in Figure 1.[22]

Our first main result is to characterize the class of OSP games and mechanisms as the class of millipede games with greedy strategies. A strategy is called *greedy* if at each move at which the agent can clinch the best still-possible outcome for her, the strategy has the agent clinch this outcome; otherwise, the agent passes.

Theorem 1. *Every obviously strategy-proof mechanism* $(\Gamma, S_{\mathcal{N}})$ *is equivalent to a millipede game with the greedy strategy. Every millipede game with the greedy strategy is obviously strategy-proof.*

This theorem is applicable in many environments. This includes allocation problems in which agents care only about the object(s) they receive, in which case, clinching actions correspond to taking a specified object and leaving the remaining objects to be distributed amongst the remaining agents. Theorem 1 also applies to standard social choice problems in which no agent is indifferent between any two outcomes (e.g., voting), in which case clinching corresponds to determining the final outcome for all agents. In such environments, Theorem 1 implies that each OSP game is equivalent to a game in which either there are only two

[21] An agent is *active* at history h of a millipede game if the agent moves at h, or the agent moved prior to history h and has not yet clinched an outcome.

[22] The first more complex example of a millipede game we know of is due to Ashlagi and Gonczarowski (2016). They construct an example of OSP-implementation of deferred acceptance on some restricted preference domains. On these restricted domains, DA reduces to a millipede game (though they do not classify the actions as "passing" or "clinching" actions). Later work by Bade and Gonczarowski (2016) gives an example of a millipede game that, while shorter, is, in some respects, even more complex than our examples.

outcomes that are possible when the first agent moves (and the first mover can either clinch any of them, or can clinch one of them or pass to a second agent, who is presented with an analogous choice, etc.), or the first agent to move can clinch any possible outcome and has no passing action. The latter case is the standard dictatorship, with a possible restricted set of possible outcomes, while the former case resembles the almost-sequential dictatorships we study in the next section.

We now outline the main ideas of the proof of Theorem 1. First, to show that greedy strategies are obviously dominant in a millipede game, note that if, at some history h, an agent can clinch her top possible outcome, it is clearly obviously dominant to do so. Harder is to show that if an agent cannot clinch her top possible outcome at h, then passing is obviously dominant. Formally, this follows from conditions (a)-(d) above, which ensure that passing will always be "safe" for her. The complete technical details can be found in the appendix.

The more difficult (and interesting) part of Theorem 1 is the first part: for any OSP game Γ, we can find an equivalent millipede game with greedy strategies. The proof in the appendix breaks the argument down into three main steps.

Step 1. *Every OSP game* Γ *is equivalent to a perfect information OSP game* Γ' *in which Nature moves once, as the first mover.*

Intuitively, this follows because if we break any information set with imperfect information to several different information sets with perfect information, the set of outcomes that are possible shrinks. For an action a to be obviously dominant, the worst possible outcome from a must be (weakly) better than the best possible outcome from any other a'. If the set of possibilities shrinks, then the worst case from a only improves, and the best case from a' worsens; thus, if a was obviously dominant in Γ, it will remain so in Γ'.[23]

Step 2. *At every history, all actions except for possibly one are clinching actions.*

Step 2 allows us to greatly simplify the class of OSP games to "clinch or pass" games. Indeed, if there were two passing actions a and a', then following each of a and a' there are at least two outcomes that are possible for i. Thus, we will always be able to find a type of agent i for which one of the possibilities following a is at best his second choice, while one of the possibilities following a' is his first choice, which implies that a does not obviously dominate a'. An equivalent argument shows that there is a type such that a' does not obviously dominate a.

[23]That every OSP game is equivalent to an OSP game with perfect information was first pointed out in a footnote by Ashlagi and Gonczarowski (2016). The same footnote also states that de-randomizing an OSP game leads to an OSP game. For completeness, the appendix contains the (straightforward) proofs of these statements.

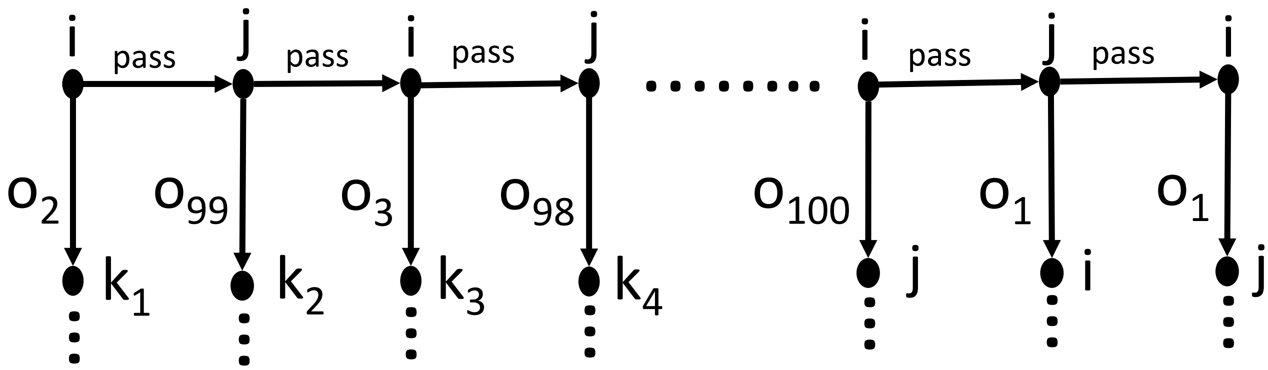

Figure 2: An example of a millipede game.

Step 3. *If agent i passes at a history h, then the payoff she ultimately receives must be at least as good as any of the payoffs she could have clinched at h.*

An agent may follow the passing action if she cannot clinch her favorite possible outcome today, and so she passes, hoping she will be able to move again in the future and get it then. To retain obvious strategy-proofness, the game needs to promise agent i that she can never be made worse off by passing. This implies that one of the conditions (a)-(d) in definition above must obtain. Combining steps 1-3 imply that any OSP game Γ is equivalent to a millipede game.

Theorem 1 characterizes the entire class of obviously strategy-proof games. We have already briefly mentioned some familiar dictatorship-like games that fit into this class (e.g., Random Priority, also known as Random Serial Dictatorship, in an object allocation environment). Another example of a millipede game is given in Figure 2. Here, there are 100 agents $\{i, j, k_1, \ldots, k_{98}\}$ and 100 objects $\{o_1, o_2, \ldots, o_{100}\}$ to be assigned. The game begins with agent i being offered the opportunity to clinch o_2, or pass to j. Agent j can then either clinch o_{99}, in which case the next mover is k_2, or pass back to i, and so on. Now, consider the type of agent i that prefers the objects in the order of their index: $o_1 \succ_i o_2 \succ_i \cdots \succ_i o_{100}$. At the very first move of the game, i is offered her second-favorite object, o_2, even though her top choice, o_1, is still available. The obviously dominant strategy here requires i to pass. However, if she passes, she may not be offered the opportunity to clinch her top object(s) for hundreds of moves. Further, when considering all of the possible moves of the other agents, if i passes, the game has the potential to go off into thousands of different directions, and in many of them, she will never be able to clinch better than o_2. Thus, while passing is formally obviously dominant, fully comprehending this still requires the ability to reason far into the future of the game and perform lengthy backwards induction.

4 Strong Obvious Strategy-Proofness

The upshot of the previous section is that some OSP mechanisms, such as Random Priority, are indeed quite simple to play; however, the full class of millipede games is much larger, and contains OSP mechanisms that may be quite complex to actually play. The reason is that OSP relaxes the assumption that agents fully comprehend how the choices of other agents will translate into outcomes, but it still presumes that they understand how their *own* future actions affect outcomes. Thus, while OSP guarantees that when taking an action, agents do not have to reason carefully about what their opponents will do, it still may require that they reason carefully about the continuation game, in particular with regard to their own "future self". Here, we introduce a strengthening of obvious dominance that we call *strong obvious dominance.*[24]

Definition 1. For an agent i with preferences $\succ_i$, strategy S_i *strongly obviously dominates* strategy S_i' in game Γ if, starting at any earliest point of departure $\mathcal{I}$ between S_i and S_i', the best possible outcome from following S_i' at $\mathcal{I}$ is weakly worse than the worst possible outcome from following S_i at $\mathcal{I}$, where the best and worst cases are determined by considering any future play by other agents (including Nature) and agent i. If a strategy S_i strongly obviously dominates all other S_i', then we say that S_i is *strongly obviously dominant.*

If a mechanism admits a profile of strongly obviously dominant strategies, we say that it is *strongly obviously strategy-proof (SOSP)*. Returning to the examples at the end of the previous section, Random Priority is SOSP, but the millipede game depicted in Figure 2 is not. Thus, strongly OSP mechanisms further delineate the class of games that are simple to play, by eliminating the more complex millipede games that may require significant forward-looking behavior and backward induction.

Strong obvious strategy-proofness has several appealing features that capture the idea of a game being simple to play. As mentioned above, SOSP strengthens OSP by looking at the worst/best case outcomes for i over all possible future actions that could be taken by i's opponents and agent i herself. Thus, a strongly obviously dominant strategy is one that is weakly better than all alternative strategies even if the agent is concerned that she might tremble in the future or has time-inconsistent preferences. We will see that SOSP games are also those that query each individual agent at most once: this ensures that the agent will not be required to look far into the future and perform lengthy backwards induction, a process which can be difficult for people in the real-world.

[24]Recall also the example of chess discussed in the introduction: if White can force a win, then any winning strategy of White is obviously dominant, yet the strategic choices in chess seem far from obvious. Chess will not admit a strongly obviously dominant strategy.

This previous discussion can be made more formal. In particular, we show that strongly obviously dominant strategies are those that are robust to agents' potential misunderstandings about the game they are playing.[25] Consider an agent who, when she is called to play at some history, knows the outcomes that are possible from each action, but may not know precisely how those outcomes depend on the future play. Strongly obviously dominant strategies are those that remain weakly dominant for such agents. To state the result formally, we consider perfect-information games (this restriction is for simplicity only). We say that an extensive-form game Γ is *outcome-set equivalent* to extensive-form game Γ' if there is a bijection ϕ between histories such that for all h, the set of possible outcomes following history h in Γ is equivalent to the set of possible outcomes following history $\phi(h)$ in Γ'. We then obtain the following result.[26]

Theorem 2. *For all i and $\succ_i$, strategy S_i is strongly obviously dominant in Γ if and only if in all outcome-set equivalent games Γ', the corresponding strategy S_i' is weakly dominant.*

Just as Theorem 1 did for OSP, we can characterize the entire class of SOSP mechanisms. Strengthening OSP to SOSP eliminates the complex examples of millipede games, such as those in Figure 2. We say that a mechanism is a *curated dictatorship* if it is a perfect-information game in which Nature moves first (if at all), and then the agents move in turn, with each agent moving at most once. The ordering of the agents and the sets of payoffs from which they choose is determined by Nature's move, and the moves of earlier agents. As long as there are at least three payoffs possible for an agent who moves, at his move, this agent can clinch any of the possible payoffs, and while clinching any of the payoffs the agents also picks a message from a pre-determined set of messages. When only two payoffs are possible, the agent can be faced with either a choice between them (including picking an accompanying message), or, he might be given a possibility to clinch one of these payoffs (and picking an accompanying message) and passing (with no message).[27]

Theorem 3. *Every strongly obviously strategy-proof mechanism $(\Gamma, S_{\mathcal{N}})$ is equivalent to a curated dictatorship with the greedy strategy. Every curated dictatorship with the greedy strategy is strongly obviously strategy-proof.*

[25] Li (2017) shows a similar result for OSP. He also shows that OSP mechanisms are precisely the mechanisms that can be implemented with bilateral commitment; this result does not extend to our setting. The reason is that bilateral commitment presumes that agents are perfectly forward looking and do not make errors in single-agent games, and SOSP relaxes this assumption.

[26] We note also that this result continues to hold even if our richness assumption on preferences is violated.

[27] Recall that in a millipede game, there may be several ways for an agent to clinch the same payoff; sending a message is a simple way to encode which of the clinching actions the agent takes. Fixing a clinched payoff, each possible message may affect the future of the game, e.g., by determining who the next mover is, though the agent's own payoff is not affected. In the literature, this property is sometimes called *bossiness*.

That curated dictatorships are SOSP is immediate; the other part of this theorem can be derived relatively easily from Theorem 1. Since SOSP implies OSP, any SOSP game is equivalent to a millipede game. Given the definition of a curated dictatorship, it is sufficient to show that pruned millipede games in which there is a history h at which the acting agent has three or more possible payoffs and a passing move are not SOSP.[28] Consider such a game and let h be any earliest history where the acting agent i has three or more possible payoffs and a passing action. The fact that h is the earliest such history implies that h is the first time i is called to act; otherwise, she must have passed at some $h' \subset h$, and, since everything that is possible at h must also be possible at any earlier h', this contradicts that h is the earliest history that satisfies our requirements. Since it is the first time i is called to move, h is on the path of play for all types of agent i. Because there can be at most one passing action, at least one payoff must be clinchable for i, say x, and at least one must not be clinchable at h, say y. Let $z \neq x, y$ be some third payoff that is possible at h. To complete the argument, note that agent i of type $y \succ_i x \succ_i z \succ_i \cdots$ has no strongly obviously dominant action at h, and so the game is not SOSP.

5 Random Priority

Thus far, all of our results have been about incentives and what makes a game "simple to play". While incentives are undoubtedly important, they are not the only consideration when designing a mechanism; efficiency and fairness are two other key goals. As an application of our new definition of SOSP, we can use it, combined with very natural fairness and efficiency axioms, to characterize the popular Random Priority (RP) mechanism.

While we have informally been using object allocation as a running example for our previous results, we now make this more concrete. There is a set of agents $\mathcal{N}$ and objects $\mathcal{O}$; each agent is to be assigned exactly one of objects, and we assume $|\mathcal{N}| = |\mathcal{O}|$. In this context, an "outcome" is an allocation of objects to agents, and the outcome space $\mathcal{X}$ is the space of all such allocations. Agents care only about their own assignment, and are indifferent between allocations where the assignment of others may vary, but their own assignment is unchanged. We consider single-unit demand for simplicity and comparison with the prior literature that often focuses on this case (Abdulkadiroğlu and Sönmez, 1998; Bogomolnaia and Moulin, 2001; Che and Kojima, 2009; Liu and Pycia, 2013), but equivalent results hold in more general models that lie in the framework laid out in Section 2.

[28] A pruned game is one for which all histories are on the path of play for some profile of types $\succ_{\mathcal{N}}$. By the pruning principle of Li (2017), it is without loss of generality to restrict to pruned games. See the proof of Theorem 1 in the appendix for more detail.

We begin by considering only incentives and efficiency, and characterize SOSP and efficient mechanisms as the class of almost-sequential dictatorships.[29] An *almost-sequential dictatorship* is a perfect-information game in which the agents are ordered and move in turn, with each agent moving at most once. At his move, an agent picks his objects and sends a message.[30] As long as there are at least three objects unallocated, the moving agent can choose from all still available objects. When there are two objects remaining, an agent can be faced with either a choice between them, or, he might be given a choice between one of these objects for sure or giving the next agent an opportunity to allocate the remaining two objects among the two of them.

The difference between a serial dictatorship and a sequential dictatorship is that in a serial dictatorship, the ordering of the agents is fixed in advance, while in a sequential dictatorship, the next agent to move may depend on the choices of the earlier agents. The reason we call the mechanism described above an "almost" sequential dictatorship is because it works exactly as a sequential dictatorship if there are three or more objects remaining, with a small modification allowed when only two objects remain.

Theorem 4. *Every strongly obviously strategy-proof and Pareto efficient mechanism* $(\Gamma, S_{\mathcal{N}})$ *is equivalent to an almost-sequential dictatorship with the greedy strategy. Every almost-sequential dictatorship with the greedy strategy is strongly obviously strategy-proof and Pareto efficient.*

Almost-sequential dictatorships satisfy two of our three desiderata (SOSP and efficiency), but they are not fair. Indeed, consider a simple serial dictatorship with a fixed ordering of the agents (which is a special case of an almost-sequential dictatorship). Then, the first agent in the ordering, i_1, always gets his first choice, while the last agent, i_N, always gets whatever remains.

Formally, a mechanism satisfies *equal treatment of equals* if, whenever two agents have the same type, they receive the same distribution over payoffs. It is easy to see that the above serial dictatorship fails this criterion. The natural way to resolve this unfairness is to order the agents randomly, and allow them to pick in this order, which gives the popular Random Priority (RP) mechanism. It is simple to check that RP is SOSP, efficient, and satisfies equal treatment of equals. The next result says that RP is in fact the *only* such mechanism.

[29] Pycia and Ünver (2016) use the same name for deterministic mechanisms without messages that belong to the class we study; they show that these are exactly the deterministic mechanisms which are strategy-proof and Arrovian efficient with respect to a complete social welfare function. We use the name their introduced because our class is a natural extension of theirs.

[30] Recall the discussion of messages in footnote 27.

Theorem 5. *Every mechanism that is strongly obviously strategy-proof, Pareto efficient, and satisfies equal treatment of equals is equivalent to Random Priority with greedy strategies. Furthermore, Random Priority with greedy strategies is strongly obviously strategy-proof, Pareto efficient, and satisfies equal treatment of equals.*

Random Priority succeeds on all three important dimensions: it is simple to play, Pareto efficient, and fair. However, this is only a partial explanation of its success, as to now, it has remained unknown whether there exist other such mechanisms, and, if so, what explains the relative popularity of RP over these alternatives.[31] Theorem 5 provides an answer to this question: not only does RP have good efficiency, fairness, and incentive properties, it is indeed the *only* mechanism that does so, thus explaining the widespread popularity of RP in practice.

6 Conclusion

In general social choice environments without transfers, we study the question of what makes a game "simple to play". We first characterize the entire class of obviously strategy-proof games, and show that they take the form of clinch-or-pass games that we call millipede games. In a millipede game, at each move, an agent is offered (potentially several) outcomes that she can clinch immediately and leave the game; in addition, she may be offered the opportunity to pass (and stay in the game). While obviously strategy-proof, some millipede games may still require extensive foresight and backwards induction, and so may not necessarily be simple to play for real-world agents. We thus propose a new definition of strong obvious strategy-proofness. Strongly obviously strategy-proof mechanisms eliminate the need to perform backwards induction, and thus are simple to play. Indeed, one strongly obviously dominant strategy-proof mechanism is seen extensively in practice: Random Priority. We use strong obvious strategy-proofness together with Pareto efficiency and equal treatment of equals to characterize Random Priority. Thus, our results show that Random Priority is the unique mechanism that has good efficiency, fairness, and incentive properties, providing an explanation for its widespread use.

[31] Bogomolnaia and Moulin (2001) provide a characterization of RP in the special case of $|\mathcal{N}| = 3$, but their result does not extend to larger markets; Liu and Pycia (2013) provide a characterization using asymptotic versions of standard axioms in replica economies as the market size grows to infinity.

A Proofs

A.1 Proof of Theorem 1

Before proceeding with the proof of Theorem 1, we first define the concepts of possible, guaranteeable, and clinchable outcomes/actions more formally.

A.1.1 Definitions

Fix a game Γ. Let $S = (S_i)_{i \in \mathcal{N}}$ denote a strategy profile for the agents, and recall that ω denotes one particular realization of Nature's moves (i.e., at each h at which nature is called to play, $\omega(h)$ is a degenerate distribution that puts all probability on exactly one action). Define $z(h, S, \omega) \in \mathcal{X}$ as the unique final outcome reached when play starts at some history h and proceeds according to (S, ω).

We first discuss the distinction between types of payoffs (possible vs. guaranteeable) and then the distinction between types of actions (clinching actions vs. passing actions). Recall that agents may be indifferent between several outcomes. For any outcome $x \in \mathcal{X}$, let $[x]_i = \{y \in \mathcal{X} : y \sim_i x\}$ denote the x-indifference class of agent i, and define

$$X_i(h, S_i) = \{[x]_i : z(h, (S_i, S_{-i}), \omega) \in [x]_i \text{ for some } (S_{-i}, \omega)\}$$

to be the possible indifference classes that may obtain for agent i starting at history h if she follows strategy S_i. Consider an agent i of type $\succ_i$. If there exists some S_i such that $[x]_i \in X_i(h, S_i)$, then we then we say that $[x]_i$ is **possible** for i at h. If, further, there exists some S_i such that $X_i(h, S_i) = \{[x]_i\}$, then we say $[x]_i$ is **guaranteeable** for i at h. Let

$$P_i(h) = \{[x]_i : \exists S_i \text{ s.t. } [x]_i \in X_i\{h, S_i)\}$$
$$G_i(h) = \{[x]_i : \exists S_i \text{ s.t. } X_i(h, S_i) = \{[x]_i\}\}$$

be the sets of payoffs that are possible and guaranteeable at h, respectively.[32] Note that $G_i(h) \subseteq P_i(h)$, and the set $P_i(h) \setminus G_i(h)$ is the set of payoffs that are possible at h, but are not guaranteeable at h. (In the proof of the theorem below, we will generally drop the bracket notation $[x]_i$ and, when there is no confusion, simply refer to the "payoff x". Statements such as "x is a possible payoff at h" or "$x \in P_i(h)$" are understood as "some outcome in the indifference class $[x]_i$ is possible at h".)

Last, we define a distinction between two kinds of actions: clinching actions and passing actions. Let i be the agent who is to act at a history h. Using our notational convention

[32] Note that $P_i(h)$ and $G_i(h)$ are well-defined even if i is not the agent who moves at h.

that (h,a) denotes the history obtained by starting at h and following action a, the set $P_i((h,a))$ is the set of payoffs that would be possible for i if she were to follow action a at h. If $P_i((h,a)) = \{[x]_i\}$, then we say that action $a \in A(h)$ **clinches** payoff x for i. If an action a clinches x for i and i never moves again following a, we call a a **clinching** action. Note that there can be more than one action that clinches the same payoff x for i, though different choices may lead to different payoffs for other agents. Any action of an agent that is not a clinching action is called a **passing** action.

We let $C_i(h)$ denote the set of payoffs that are clinchable for i at h.[33] In words, following a clinching action, i's outcome is completely determined (modulo indifference classes), and i is never called on to move again. Last, note that this definition of $C_i(h)$ presumes that agent i is called to play at history h. If $\bar{h}$ is a terminal history, then no agent is called to play and there are no actions. However, it will be useful in what follows to define $C_i(\bar{h}) = \{[x]_i\}$ for all i, where x is the unique outcome associated with the terminal history $\bar{h}$.

A.1.2 Proof

With the above definitions in hand, we can prove Theorem 1. We start by proving that millipede games are OSP (Proposition 1), and then prove that every OSP game is equivalent to a millipede game (Proposition 2).

Proposition 1. *Millipede games with greedy strategies are obviously strategy-proof.*

Proof. Let Γ be a millipede game. Recall that the *greedy strategy* for any agent i is defined as follows: for any history h at which i movies, if i can clinch her top payoff in $P_i(h)$, then $S_i(\succ_i)(h)$ instructs i to follow an action that clinches this payoff; otherwise, i passes at h.[34]

We now show that it is obviously dominant for all agents to follow a greedy strategy. Consider some profile of greedy strategies $(S_i(\cdot))_{i\in\mathcal{N}}$. For any subset of outcomes $X' \subset \mathcal{X}$, define $Top(\succ_i, X')$ as the best possible payoff in the set X' according to preferences $\succ_i$, i.e., $x \in Top(\succ_i, X')$ if and only if $x \succsim_i y$ for all $y \in X'$ (note that we use our standard convention whereby a payoff x represents the entire indifference class to which x belongs, and so $Top(\succ_i, X')$ is effectively unique). Then, $Top(\succ_i, P_i(h))$ denotes i's top payoff among all payoffs that are possible at history h, and $Top(\succ_i, C_i(h))$ denotes i's top payoff among all of his clinchable payoffs at h. It is clear that if $Top(\succ_i, C_i(h)) = Top(\succ_i, P_i(h))$, then the greedy action of clinching the top payoff is obviously dominant at h. What remains to be shown is if $Top(\succ_i, C_i(h)) \neq Top(\succ_i, P_i(h))$, then passing is obviously dominant at h.

[33] That is, $C_i(h) = \{[x]_i : \exists a \in A(h) \text{ s.t. } P_i((h,a)) = \{[x]_i\}\}$.

[34] There may be multiple ways for i to clinch the same payoff x at h, and further, x may in principle still be possible/guaranteeable if i passes at h. Our goal is simply to prove the existence of at least one obviously dominant strategy for i.

Assume that there exists a history h that is on the path of play for type $\succ_i$ when she follows the greedy strategy and $Top(\succ_i, C_i(h)) \neq Top(\succ_i, P_i(h))$, yet passing is not obviously dominant at h; further, let h be any earliest such history for which this is true. To shorten notation, let $x_P(h) = Top(\succ_i, P_i(h))$, $x_C(h) = Top(\succ_i, C_i(h))$, and let $x_W(h)$ be the worst possible outcome from passing (and following the greedy strategy in the future). Since passing is not obviously dominant, it must be that $x_W(h) \not\succsim_i x_C(h)$.

First, note that $h' \in \mathcal{H}_i(h)$ implies $x_W(h) \succsim_i x_W(h')$. Since passing is obviously dominant at all $h' \in \mathcal{H}_i(h)$, we have $x_W(h') \succsim_i x_C(h')$, and together, these imply that $x_W(h) \succsim_i x_C(h')$ for all $h' \in \mathcal{H}_i(h)$. At h, since passing is not obviously dominant, we have $x_C(h) \succ_i x_W(h)$, and further, there must be some $x' \in P_i(h) \setminus G_i(h)$ such that $x' \succ_i x_C(h) \succ_i x_W(h)$.[35] The above implies that $x' \succ_i x_C(h) \succ_i x_C(h')$ for all $h' \in \mathcal{H}_i(h)$. Let $X_0 = \{x' : x' \in P_i(h) \text{ and } x' \succ_i x_C(h)\}$. In words, X_0 is a set of payoffs that are possible at all $h' \subseteq h$, and are strictly better than anything that was clinchable at any $h' \subseteq h$ (and therefore have never been clinchable themselves). Order the elements in X_0 according to $\succ_i$, and wlog, let $x_1 \succ_i x_2 \succ_i \cdots \succ_i x_M$.

Definition. Let h' be a history where either agent i is called to move or h' is a terminal history. Payoff x **becomes impossible** for i at h' if: (i) $x \in P_i(h'')$ for all $h'' \in \mathcal{H}_i(h')$ and (ii) $x \notin P_i(h')$.

In other words, the history h' at which a payoff becomes impossible is the earliest history at which i moves and where x "disappears" as a possible outcome for her.

Consider a path of play starting from h and ending in a terminal history $\bar{h}$ at which type $\succ_i$ of agent i receives his worst case outcome $x_W(h)$. For every $x_m \in X_0$, let h_m denote the history on this path at which x_m becomes impossible for i. Note that because i is ultimately receiving payoff $x_W(h)$, such a history h_m exists for all $x_m \in X_0$.[36] Let $\hat{h} = \max\{h_1, h_2, \ldots, h_M\}$ (ordered by $\subset$); in words, $\hat{h}$ is the earliest history at which everything in X_0 is no longer possible. Further, let $\hat{h}_{-m} = \max\{h_1, \ldots, h_{m-1}\}$, i.e., $\hat{h}_{-m}$ is the earliest history at which all payoffs strictly preferred to x_m are no longer possible.

Claim 1. For all $x_m \in X_0$ and all $h' \subseteq \bar{h}$, we have $x_m \notin C_i(h')$.

Proof. First, note that $x_m \notin C_i(h')$ for any $h' \subseteq h$ by construction. We will show that $x_m \notin C_i(h')$ at any $\bar{h} \supseteq h' \supset h$ as well. Start by considering $m = 1$, and assume $x_1 \in C_i(h')$ for some $\bar{h} \supseteq h' \supset h$. By definition, $x_1 = Top(\succ_i, P_i(h))$; since $h' \supset h$ implies that $P_i(h') \subseteq$

[35] At least one such x' exists by the assumption that $Top(\succ_i, C_i(h)) \neq Top(\succ_i, P_i(h))$, though there in general may be multiple such x'.

[36] It is possible that h_m is a terminal history.

$P_i(h)$, we have that $x_1 = Top(\succ_i, P_i(h'))$ as well. Since $x_1 \in C_i(h')$ by supposition, greedy strategies direct i to clinch x_1, which contradicts that she receives $x_W(h)$.[37]

Now, consider an arbitrary m, and assume that for all $m' = 1, \ldots, m-1$, payoff $x_{m'}$ is not clinchable at any $h' \subseteq \bar{h}$, but x_m is clinchable at some $h' \subseteq \bar{h}$. Let $x_{m'}$ be (a) payoff that becomes impossible at $\hat{h}_{-m}$ and is such that $x_{m'} \succ_i x_m$. There are two cases:

Case (i): $h' \subset \hat{h}_{-m}$**.** This is the case where x_m is clinchable while there is some strictly preferred payoff $x_{m'} \succ_i x_m$ that is still possible. Note that at history $\hat{h}_{-m}$, neither (a) nor (b) in the definition of a millipede game hold for $x_{m'}$. Further, by the inductive hypothesis, $x_{m'}$ is nowhere clinchable, and so (c) does not hold either. This implies that (d) must hold, and so x_m must be clinchable at $\hat{h}_{-m}$. Then, since all preferred payoffs are no longer possible at $\hat{h}_{-m}$, x_m is the best possible payoff remaining, and is clinchable. Therefore, greedy strategies instruct agent i to clinch x_m, which contradicts that she receives $x_W(h)$.

Case (ii): $h' \supseteq \hat{h}_{-m}$**.** In this case, x_m becomes clinchable after all strictly preferred payoffs are no longer possible. Thus, again, greedy strategies instruct i to clinch x_m, which contradicts that she is receiving $x_W(h)$. □

To finish the proof, again let $\hat{h} = \max\{h_1, h_2, \ldots, h_M\}$ and let $\hat{x}$ be a payoff that becomes impossible at $\hat{h}$. The claim shows that $\hat{x}$ is not clinchable at any $h' \subseteq \hat{h}$. The preceding two statements imply that conditions (a)-(c) in the definition of a millipede game do not hold at $\hat{h}$, and so condition (d) must hold, i.e., $x_C(h) \in C_i(\hat{h})$. Since $x_C(h)$ is the best possible remaining payoff at $\hat{h}$, greedy strategies direct i to clinch $x_C(h)$, which contradicts that she receives $x_W(h)$.[38] □

We now prove the second part of the theorem, restated below as Proposition 2. To do so, we first need to introduce the pruning principle of Li (2017), which will simplify some of the arguments. Given a game Γ and strategy profile $(S_i(\succ_i))_{i\in\mathcal{N}}$, the **pruning** of Γ with respect to $(S_i(\succ_i))_{i\in\mathcal{N}}$ is a game Γ' that is defined by starting with Γ and deleting all histories of Γ that are never reached for any type profile. Then, the **pruning principle** says that if $(S_i(\succ_i))_{i\in\mathcal{N}}$ is obviously dominant for Γ, the restriction of $(S_i(\succ_i))_{i\in\mathcal{N}}$ to Γ' is obviously dominant for Γ', and both games result in the same outcome. Thus, for any OSP mechanism, we can find an equivalent OSP pruned mechanism. When proving this proposition, we assume that all OSP games have been pruned with respect to the equilibrium strategy profile. Note also that we actually prove a slightly stronger statement, which is that every OSP game is

[37] Recall that for terminal histories h, we define $C_i(h) = \{x\}$, where x is the unique payoff associated with the terminal history. Thus, if h' is a terminal history, then i receives payoff x_1, which also contradicts that she receives payoff $x_W(h)$.

[38] If $\hat{h}$ is a terminal history, then we make an argument analogous to footnote 37 to reach the same contradiction.

equivalent to a millipede game that satisfies the following additional property: for all i, all h at which i moves, and all $x \in G_i(h)$, there exists an action $a_x \in A(h)$ that clinches x.[39]

Proposition 2. *Every obviously strategy-proof mechanism* $(\Gamma, S_{\mathcal{N}})$ *is equivalent to a millipede game with the greedy strategy.*

Proof. The proof of this proposition is broken down into several lemmas.

Lemma 1. *Every OSP game is equivalent to an OSP game with perfect information in which Nature moves at most once, as the first mover.*

Proof. Ashlagi and Gonczarowski (2016) briefly mention this result in a footnote; here, we provide the straightforward proof for completeness. We first show that every OSP game is equivalent to an OSP game with perfect information. Denote by $A(\mathcal{I})$ the set of actions available at information set $\mathcal{I}$ to the agent who moves at $\mathcal{I}$. Take an obviously strategy-proof game Γ and consider its perfect-information counterpart Γ', that is the perfect information game at which at every history h in Γ the moving agent's information set is $\{h\}$ in Γ', the available actions are $A(\mathcal{I})$, and the outcomes in Γ' following any terminal history are the same as in Γ. Notice that the support of possible outcomes at any history h in Γ' is a subset of the support of possible outcomes at $\mathcal{I}(h)$ in Γ. Thus, the worst-case outcome from any action (weakly) increases in Γ', while the best-case outcome (weakly) decreases. Thus, if there is an obviously dominant strategy in Γ, following the analogous strategy in Γ' continues to be obviously dominant. Hence, Γ' is obviously strategy-proof and equivalent to Γ.

We now show that every OSP game is equivalent to a perfect-information OSP game in which Nature moves once, as the first mover. Consider a game Γ, which, by the previous paragraph, we can assume has perfect information. Let $\mathcal{H}_{\text{nature}}$ be the set of histories h at which Nature moves in Γ. Consider a modified game Γ' in which at the empty history Nature chooses actions from $\times_{h\in\mathcal{H}_{\text{nature}}} A(h)$. After each of Nature's initial moves, we replicate the original game, except at each history h at which Nature is called to play, we delete Nature's move and continue with the subgame corresponding to the action Nature chose from $A(h)$ at $\emptyset$. Again, note that for any agent i and history h at which i is called to act, the support of possible outcomes at h in Γ' is a subset of the support of possible outcomes at the corresponding history in Γ (where the corresponding histories are defined by mapping the $A(h)$ component of the action taken at $\emptyset$ by Nature in Γ' as an action made by Nature at h in game Γ). Using reasoning similar to the previous paragraph, we conclude that Γ' is obviously strategy-proof, and Γ and Γ' are equivalent. $\square$

[39] See Lemma 3 below.

Lemma 2. *Let Γ be an obviously strategy-proof game of perfect information that is pruned with respect to the obviously dominant strategy profile $(S_i(\succ_i))_{i\in\mathcal{N}}$. At any history h where an agent i is called to move, there is at most one action $a^* \in A(h)$ such that $P_i((h,a^*)) \nsubseteq G_i(h)$.*

Proof. We prove this result via two steps.

Step 1. Suppose there are two distinct payoffs $x, y \in P_i(h) \setminus G_i(h)$ and a preference type $\succ_i$ such that (i) x and y are the first and second $\succ_i$ −best possible payoffs in $P_i(h)$, and (ii) h is on the path of the game for type $\succ_i$. Then, there is at most one action $a^* \in A(h)$ such that $P_i((h,a^*)) \nsubseteq G_i(h)$, and type $\succ_i$ must choose action a^* at h.

To prove the claim of this step, it is enough to consider the case $x \succ_i y$. First, note that if $x \in P_i((h,a))$ for some $a \in A(h)$, then $y \in P_i((h,a))$ as well. Indeed, if not then $x \in P_i((h,a))$ and $y \notin P_i((h,a))$. For type $\succ_i$, the worst case outcome from following a is strictly worse than y because x and y are assumed to be the $\succ_i$ −best possible payoffs in $P_i(h)$, x is not guaranteeable at h, and y is not possible following a. Because $y \in P_i(h)$, action a is not obviously dominant. As x is not guaranteeable at h, the worst case outcome from any other $a' \neq a$ is strictly worse than x, while the best case outcome from a is x, and so no a' can obviously dominate a. Thus, type $\succ_i$ has no obviously dominant action, which contradicts that the game is OSP, and proves the claim of this paragraph.

Now, assume that there are two actions a_1^* and a_2^* such that $P_i((h,a_j^*)) \nsubseteq G_i(h)$ for $j = 1,2$. Consider some $x \in P_i(h) \setminus G_i(h)$. By the previous paragraph, we know that $x \in P_i((h,a_1^*))$ and $x \in P_i((h,a_2^*))$. However, by assumption, x is not guaranteeable at h, and so, for type $\succ_i$, the worst case payoff from *any* action a' must be strictly worse than x, while the best case outcomes from a_1^* and a_2^* are both x. Therefore, no action $a' \in A(h)$ is obviously dominant, which contradicts that Γ is OSP.[40]

We can conclude that there is at most one action a^* that leads to x for some continuation strategies of players. Because x is only possible following a^*, if an obviously dominant strategy profile exists, any type $\succ_i$ that ranks any $x \in P_i(h) \setminus G_i(h)$ first among the payoffs in $P_i(h)$ must select a^* at this history, concluding the proof of the claim of Step 1.

Step 2. Suppose i moves at history h. Then, there is at most one action $a^* \in A(h)$ such that $P_i((h,a^*)) \nsubseteq G_i(h)$.

To prove this step, first consider any earliest history h_0^i at which i is to move. Note that since this is an earliest history for i, history h_0^i is on the path of play for all types of agent i. If $P_i(h_0^i) \setminus G_i(h_0^i) = \{x\}$ and there were two actions a_1^* and a_2^* as in the statement, then for any type that ranks x first, the worst case from any action is strictly worse than x

[40]Note that if there were only one such action a_1^*, then it would still be true that there is no action that obviously dominates a_1^* for this type. However, a_1^* itself might be obviously dominant. When there are two such actions, a_1^* does not obviously dominate a_2^*, nor does a_2^* obviously dominate a_1^*, and thus there are no obviously dominant actions.

(because x is not guaranteeable), while the best case from both a_1^* and a_2^* is x, so nothing can obviously dominate a_1^*; by similar reasoning, a_1^* does not obviously dominate a_2^*. If there are two payoffs $x, y \in P_i(h_0^i) \setminus G_i(h_0^i)$, then we can apply Step 1 to type $x \succ_i y \succ_i \cdots$.

Now, consider any successor history $h' \supset h_0^i$ at which i is to move, and make the inductive assumption that at every $h \subset h'$, there is only one possible action $a^* \in A(h)$ such that $P_i((h, a^*)) \not\subseteq G_i(h)$. First, consider the case where $|P_i(h') \setminus G_i(h')| = 1$, and let x be the unique payoff that is possible but not guaranteeable. By way of contradiction, assume there were two actions, a_1^* and a_2^*, such that x was a possible outcome. By the pruning principle, some type $\succ_i$ must be receiving x at some terminal history $\bar{h} \supset h'$. If x is the top choice among all payoffs in $P_i(h')$ for some type $\succ_i$ for which history h' is on-path, then, following similar reasoning as above, neither a_1^* nor a_2^* can be obviously dominant. If x is not the best possible payoff in $P_i(h')$ for any such type, then, since x is the only payoff that is possible, but not guaranteeable, every other payoff in $P_i(h')$ is guaranteeable. This implies that every type of agent i for which h' is on-path can guarantee herself her best possible outcome in $P_i(h')$, and so no type should ever play a strategy for which she ends up receiving a payoff of x at any terminal history $\bar{h} \supset h'$, which is a contradiction.

Finally, assume $|P_i(h') \setminus G_i(h')| \geq 2$, and let $x, y \in P_i(h') \setminus G_i(h')$. First, consider the case that there is some $x \in P_i(h') \setminus G_i(h')$ such that $x \notin G_i(h)$ for all $h \subset h'$ at which i is to move. Recall the inductive hypothesis says that at all such h, there is a unique action such that $P_i((h, a^*)) \not\subseteq G_i(h)$. Thus, if x has never been guaranteeable, all types $x \succ_i \cdots$ must have followed this unique action at all such $h \subset h'$, and we can apply Step 1 to the type $x \succ_i y \succ_i \cdots$. The last case to consider is where all $x, y \in P_i(h') \setminus G_i(h')$ were also guaranteeable at some earlier history h. Consider a type $\succ_i$ of agent i who receives payoff x at some terminal history $\bar{h} \supset h'$. First, note that for any $z \succ_i x$, $z \notin G_i(h')$ as otherwise this type would not follow a strategy whereby x was a possible outcome.[41] Let z be the $\succ_i -$best payoff in $P_i(h')$, and w be the second-best possible payoff in $P_i(h')$ for this type; we allow $w = x$. We can again apply Step 1 to type $\succ_i$ and conclude there is at most one action a^* such that $P_i((h, a^*)) \not\subseteq G_i(h)$. $\square$

Clinching actions are those for which i's payoff is completely determined after following the action. Lemma 2 shows that if a game is OSP, then at every history, for all actions a with the exception of possibly one special action a^*, all payoffs that are possible following a are also guaranteeable at h; note, however, it does *not* say that all actions but at most one are clinching actions. Indeed, it leaves open the possibility that there are several actions that can ultimately lead to multiple final payoffs for i, which can happen when different payoffs

[41] If x is the $\succ_i$-best possible payoff in $P_i(h')$ for all types that reach h', then apply the same argument to a type that receives y at some terminal history and set $z = x$.

are guaranteeable for i by following different actions in the future of the game. The next lemma shows that if this is the case, we can always construct an equivalent OSP game such that all actions except for possibly one are clinching actions.

Lemma 3. *Let Γ be an obviously strategy-proof game of perfect information that is pruned with respect to the strategy profile $(S_i(\succ_i))_{i\in\mathcal{N}}$. There exists an equivalent obviously strategy-proof game Γ' with perfect information such that:*

(i) At each history h, at least $|A(h)|-1$ actions at h are clinching actions.

(ii) For every payoff $x \in G_i(h)$, there exists an action $a_x \in A(h)$ that clinches x.

Proof. Consider some history h of game Γ at which i moves. By Lemma 2, all but at most one action (denoted a^*) in $A(h)$ satisfy $P_i((h,a)) \subseteq G_i(h)$; this means that any obviously dominant strategy for type $\succ_i$ that does not choose a^* guarantees the best possible outcome in $P_i(h)$ for type $\succ_i$. Thus the set $\mathcal{S}_i(h) = \{S_i : |X(h,S_i)| = 1\}$ contains all possible obviously dominant strategies of agent i for which h is on path and that do not choose a^*. Notice that $\mathcal{S}_i(h)$ is the set of strategies that guarantee some payoff x for i if i plays strategy S_i starting from history h.

We create a new game Γ' that is the same as Γ, except we replace the subgame starting from history h with a new subgame defined as follows. If there is an action a^* such that $P_i((h,a^*)) \nsubseteq G_i(h)$ in the original game (of which there can be at most one), then there is an analogous action a^* in the new game, and the subgame following a^* is exactly the same as in the original game Γ. Additionally, there are $M = |\mathcal{S}_i(h)|$ other actions at h, denoted $a_1,\ldots,a_M$. Each a_m corresponds to one strategy $S_i^m \in \mathcal{S}_i(h)$, and following each a_m, we replicate the original game, except that at any future history $h' \supseteq h$ at which i is called on to act, all actions (and their subgames) are deleted and replaced with the subgame starting with the action $a' = S_i^m(h')$ that i would have played in the original game had she followed strategy $S_i^m(\cdot)$. In other words, if i were to choose some action $a \neq a^*$ at h in the original game, then, in the new game Γ', we ask agent i to choose not only her current action, but all future actions that she would have chosen according to $S_i^m(\cdot)$ as well. By doing so, we have created a new game in which every action (except for a^*, if it exists) at h clinches some payoff x, and further, agent i is never called upon to move again.[42]

We construct strategies in Γ' that are the counterparts of strategies from Γ, so that for all agents $j \neq i$, they continue to follow the same action at every history as they did in the original game, and for i, at history h in the new game, she takes the action a_m that is

[42]More precisely, all of i's future moves are trivial moves in which she has only one possible action; hence these histories may further be removed to create an equivalent game in which i is never called on to move again. Note that this only applies to the actions $a \neq a^*$; it is still possible for i to follow a^* at h and be called upon to make a non-trivial move again later in the game.

associated with the strategy S_i^m in the original game. By definition if all the agents follow strategies in the new game analogous to the their strategies from the original game, the same terminal history will be reached, and so Γ and Γ' are equivalent under their respective strategy profiles.

We must also show that if a strategy profile is obviously dominant for Γ, this modified strategy profile is obviously dominant for Γ'. To see why the modified strategy profile is obviously dominant for i, note that if her obviously dominant action in the original game was part of a strategy that guarantees some payoff x, she now is able to clinch x immediately, which is clearly obviously dominant; if her obviously dominant strategy was to follow a strategy that did not guarantee some payoff x at h, this strategy must have directed i to follow a^* at h. However, in Γ', the subgame following a^* is unchanged relative to Γ, and so i is able to perfectly replicate this strategy, which obviously dominates following any of the clinching actions at h in Γ'. In addition, the game is also obviously strategy-proof for all $j \neq i$ because, prior to h, the set of possible payoffs for j is unchanged, while for any history succeeding h where j is to move, having i make all of her choices earlier in the game only shrinks the set of possible outcomes for j, in the set inclusion sense. When the set of possible outcomes shrinks, the best possible payoff from any given strategy only decreases (according to j's preferences) and the worst possible payoff only increases, and so, if a strategy was obviously dominant in the original game, it will continue to be so in the new game. Repeating this process for every history h, we are left with a new game where, at each history, there are only clinching actions plus (possibly) one passing action, and further, every payoff that is guaranteeable at h is also clinchable at h. □

Lemma 4. *Let Γ be an obviously strategy-proof game that is pruned with respect to the obviously dominant strategy profile $S_{\mathcal{N}}$ and that satisfies Lemmas 1 and 3. At every history h^i at which an agent i moves and every terminal history $\bar{h}$, for every payoff z, one of conditions (a)-(d) must hold.*

Proof. Assume not. First, consider the case of a non-terminal history h^i where i is called to move and a payoff z be such that (a), (b) and (c) do not hold at h^i, i.e., the following are true:

(a') $z \notin P_i(h^i)$

(b') $z \in P_i(h)$ for all $h \in \mathcal{H}_i(h^i)$

(c') $z \notin \cup_{\tilde{h} \in \mathcal{H}_i(h^i)} C_i(\tilde{h})$ and

Points (b') and (c') imply that z is possible at every $h \subsetneq h^i$ where i is to move, but it is not clinchable at any of them. This implies that for any type of agent i that ranks z first, any obviously dominant strategy must have the agent passing at all $h \in \mathcal{H}_i(h^i)$.[43]

Towards a contradiction, assume that (d) did not hold, i.e., there exists some $h' \in \mathcal{H}_i(h^i)$ and $x \in C_i(h')$ such that $x \notin C_i(h^i)$. Consider a type $z \succ_i x \succ_i \cdots$. We argue that if (d) does not hold at h^i, then there is some $\hat{h}^i \subsetneq h^i$ such that type $\succ_i$ has no obviously dominant action. First, note that at any such $\hat{h}^i \subsetneq h^i$, no clinching action can be obviously dominant, because z is always possible following the passing action, but is never clinchable, and so the worst case from clinching is strictly worse than the best case from passing, which is z. Next, there must be some $\hat{h}^i \subsetneq h^i$ such that the passing action also is not obviously dominant. To see why, note that h^i must be on the path of play for type $\succ_i$, since she must pass at all $h' \subsetneq h^i$. By assumption, $z \notin P_i(h^i)$ and $x \notin C_i(h^i)$, which implies that the worst case outcome from passing at any $h' \subsetneq h^i$ is some y that is strictly worse than x according to $\succ_i$. However, we also have $x \in C_i(\hat{h}^i)$ for some $\hat{h}^i \subsetneq h^i$, and so, the best case outcome from clinching x at $\hat{h}^i$ is x. This implies that passing is not obviously dominant, which contradicts that Γ is OSP.

Last, consider a terminal history $\bar{h}$. As above, let z be a payoff such that (a'), (b'), and (c') hold (i.e., (a), (b), and (c) are false). By definition, $C_i(\bar{h}) = \{y\}$ for all i, where y is the unique outcome associated with terminal history $\bar{h}$ (note also that $z \notin P_i(\bar{h})$ implies that $y \neq z$). Assume that (d) does not hold, i.e., there exists some $h' \in \mathcal{H}_i(\bar{h})$ and $x \in C_i(h')$ such that $x \notin C_i(\bar{h})$. Note that (i) $z \neq y$ (because $z \notin P_i(\bar{h})$); (ii) $z \neq x$ (by (c')); and (iii) $x \neq y$ (because $x \notin C_i(\bar{h})$). In other words, x, y, z must all be distinct payoffs. Consider the type $z \succ_i x \succ_i y \succ_i \cdots$. By (b') and (c'), z is possible at every $h \subsetneq \bar{h}$ where i is to move, but is not clinchable at any such history. Thus, any obviously dominant strategy of type $\succ_i$ must have agent i passing at any such history. However, at h', i could have clinched x, and so passing is not obviously dominant (because y is possible from passing). □

Proposition 2 follows from Lemmas 1, 3, and 4. Theorem 1 then follows from Propositions 1 and 2.

A.2 Proof of Theorem 2

Consider any two strategies S_i and S_i' of agent i; let h be the earliest point of departure for these two strategies.

[43]At all such h, since z is not clinchable, but is possible, it must be possible following the (unique) passing action. This means that best case outcome from passing is z, while the worst case outcome from clinching is strictly worse than z. Thus, no clinching action can be obviously dominant, so, if an obviously dominant strategy exists, it must instruct i to pass.

Suppose S_i is strongly obviously strategy-proof in Γ. Then any outcome that is possible after playing S_i is weakly better than any outcome that is possible after playing S_i' in game Γ', and hence in the outcome set-equivalent game Γ'. Hence, S_i strongly obviously dominates S_i' in game Γ', and thus it weakly dominates it.

Now suppose S_i weakly dominates S_i' in all games Γ' that are outcome equivalent to Γ. Consider such a Γ' in which all moves of agent i following history h are made by Nature instead. Since, S_i weakly dominates S_i' in Γ', we conclude that any outcome that is possible after playing S_i is weakly better than any outcome that is possible after playing S_i' in game Γ', and hence in the outcome set-equivalent game Γ. Hence, S_i strongly obviously dominates S_i' in game Γ. □

A.3 Proof of Theorem 4

It is obvious that almost-sequential dictatorships are SOSP and efficient. For the other direction, note that by our previous Lemma 3, any OSP game is equivalent to one such that there is at most one non-clinching move at each history, and everything that is guaranteeable is also clinchable. We show that in fact, for every history that is not penultimate to a terminal history, all moves must be clinching moves. By strengthening OSP to strong OSP, following any move, we need only consider the entire set of possible outcomes for i following any action.[44]

We proceed by induction. Consider $N = 2$, and denote $\mathcal{N} = \{i, j\}$ and $\mathcal{O} = \{o_1, o_2\}$.[45] Consider any efficient and SOSP game Γ. Without loss of generality, let the first mover be i, and note that by efficiency, both o_1 and o_2 must be possible for her. Again without loss of generality, assume she can clinch o_1 at the first move (she must be able to clinch at least one of o_1 or o_2, since there can be at most one non-clinching move). Consider the first agent who is offered the opportunity to clinch o_2 (following starting the game by a series of passes). If this agent is i, then it is equivalent to offer her the opportunity to clinch o_2 at her first move, and the mechanism is again a serial dictatorship. If the first person to be able to clinch o_2 is j, then it is equivalent to offer her the opportunity to clinch o_2 at her first move, and the game is an almost sequential dictatorship.

Consider now $N = 3$, where $\mathcal{N} = \{i, j, k\}$ and $\mathcal{O} = \{o_1, o_2, o_3\}$, and let the first mover be i. By efficiency, all items are possible for her at the initial history. Assume she had a non-clinching move. This means for one of her actions, labeled a^*, there are (at least) two

[44] Under OSP, the current mover may have "veto power" over some future outcomes, but not others; however, this requires reasoning about the future, and so is eliminated by strong OSP.

[45] Recall that $\mathcal{O}$ is the set of objects to be assigned, while the outcome space $\mathcal{X}$ is the set of all possible allocations. Since an agent's payoff is determined by only her own allocation, we will use the notation $o_1 \succ_i o_2 \succ_i \cdots$ to describe an agent's type.

possible outcomes, o_1 and o_2, at least one of which (say o_1) is not clinchable at the initial history. There are two cases, depending on whether the third outcome o_3 is clinchable or not:

o_3 is clinchable at the initial history: By assumption, o_3 is clinchable and o_1 is not. Consider type $o_1 \succ_i o_3 \succ_i o_2$. None of her clinching actions are strongly obviously dominant, since o_1 is possible following a^*. In addition, a^* is also not strongly obviously dominant, since o_2 is possible, but she could have clinched o_3. Thus, this type of agent i has no strongly obviously dominant strategy.

o_3 is not clinchable at the initial history: In this case, o_2 is clinchable, but o_1 and o_3 are not (since only one passing move is allowed). Then, consider type $o_3 \succ_i o_2 \succ_i o_1$. o_3 must be possible (by efficiency), and so must be possible following a^*. This means that no clinching action is strongly obviously dominant. Following the (unique) non-clinching action is also not strongly obviously dominant, because o_1 is possible following a^*, while o_2 is clinchable.

Thus, the first agent to move must have only clinching actions, and, by efficiency, must be able to clinch any object. Following any such clinching move, the game is equivalent to a game of size $N = 2$, which we have already shown is equivalent to an almost-sequential dictatorship.

Last, assume that for every market of size $n = 1, \ldots, N-1$ any efficient and SOSP game is equivalent to an almost sequential dictatorship. Consider a market of size N. Let $\mathcal{N} = \{i_1, \ldots, i_N\}$ and $\mathcal{O} = \{o_1, \ldots, o_N\}$. By efficiency, all items are possible for the first mover, i_1, at the initial history. We argue that all of her actions must be clinching actions.

Assume not. Then there is exactly one action a^* that is a passing action. By definition of a passing action, there must be (at least) two possible outcomes, o_1 and o_2, at least one of which (o_1, say) is not clinchable at the initial history. There are two cases:

There exists a $z \neq o_1, o_2$ that is clinchable at the initial history: By assumption, z is clinchable and o_1 is not. Consider type $o_1 \succ_{i_1} z \succ_{i_1} \cdots$. None of the clinching actions are strongly obviously dominant, since o_1 is possible following a^*, but cannot be clinched. In addition, a^* is not strongly obviously dominant, because o_2 is possible following a^*, while z is clinchable.

There does not exist a $z \neq o_1, o_2$ that is clinchable at the initial history: In this case, o_2 must be clinchable, while all $z \neq o_2$ are not. Choose some $z \neq o_1, o_2$, and consider the type $z \succ_{i_1} o_2 \succ_{i_1} o_1 \succ_i \cdots$. Since z is possible (by efficiency), clinching o_2 is not strongly obviously dominant. However, since o_1 is possible following a^*, while o_2 is clinchable, a^* is not strongly obviously dominant either.

Thus, we have that the first mover, i_1, must have only clinching actions, and she must be able to clinch everything. Following any of i_1's clinching actions, we have a game of size

$N-1$, which, by the inductive hypothesis, is an almost sequential dictatorship. It is then simple to see that the overall game is also an almost sequential dictatorship.

A.4 Proof of Theorem 5

Suppose a game Γ is strongly obviously strategy-proof, efficient, and satisfies equal treatment of equals. Our characterization of SOSP and efficient mechanisms tells us that we can assume that Γ' is equivalent to an almost-sequential dictatorship, which can be run as follows: at each history, including the empty history, Nature chooses an agent from among those agents who have yet to move, and this agent moves. If there are three or more objects or exactly one object still unallocated, then this agent selects his most preferred still available object and sends an additional message. If, for the first time, there are exactly two unallocated objects (and thus two agents yet to move), then the agent who moves either (i) selects his most preferred object and sends a message, or (ii) has a choice of clinching one of the two objects (and sending a message) or passing. In the latter case, Nature selects the other agent who has yet to move, who chooses his best object, and the agent who passed obtains the remaining object.

The reminder of the proof is by induction on the number k of agents that have been called to move. We will show that each time Nature calls an agent to move, it must select uniformly at random from all agents who have yet to be called. Consider $k=1$ (i.e., the first agent called), and suppose there are more than three agents and objects. Consider a preference profile where all agents rank objects in the same order, $o_1 \succ o_2 \succ \cdots \succ o_N$. By equal treatment of equals, all agents must receive object o_1 with equal probability. Since, at this profile, o_1 is always taken by the first mover, we conclude that Nature must select each agent to be the first mover with equal probability.

Now, consider any k-th move, where $k<N$, and assume that for all moves $1,\ldots,k-1$, each remaining agent at that point was called on by Nature with equal probability. Label the history under consideration as h, and name the agents so that along the path to h, the first mover is i_1 who chooses o_1, then i_2 chooses o_2, etc. until i_{k-1} chooses o_{k-1}.

Suppose first that there are at least three objects left. We claim that each agent who has not moved yet has an equal chance to be called by Nature at history h. To see why, consider some object $o \neq o_1, ..., o_{k-1}$ and the preference profile in which each agent i_ℓ, for $\ell = 1, ..., k-1$, ranks objects so that

$$o_1 \succ_{i_\ell} o_2 \succ_{i_\ell} ... \succ_{i_\ell} o_\ell \succ_{i_\ell} o \succ_{i_\ell} \cdots$$

and other objects are ranked below o. Let $Y(h) = I - \{i_1, \ldots, i_{k-1}\}$ be the set of agents who have yet to move at h, and assume that all $i \in Y(h)$ have the same preferences and rank o first. The total probability that i receives o is equal to the sum of (1) the probability that Nature chooses i at h and (2) the probability that Nature chooses i to move at any other $h' \neq h$ where o is still available. The crux of the argument is to note that for the preferences specified, for any other branch of the tree (that does not contain history h), object o will be claimed by someone at the $(k-1)$-th move or earlier. By the inductive hypothesis, each time Nature picks an agent to move at any of these histories, all agents in $Y(h)$ have an equal probability of being picked, and, if they are, they will immediately claim object o. This implies that (2) is the same for all agents $i \in Y(h)$. Since the sum (1)+(2) must also be the same for all agents (by equal treatment of equals), we conclude that (1) is also the same for all $i, j \in Y(h)$, i.e., the probability that i is chosen to move at h is equal to the probability that j is chosen to move at h, as desired.

Last, suppose that there are exactly two objects o and o' left at history h. With two objects, Bogomolnaia and Moulin (2001) show that Random Priority is the only strategy-proof and efficient mechanism that satisfies equal treatment of equals, and the claim follows from their work.[46]

References

Abdulkadiroğlu, A. and T. Sönmez (1998): "Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems," *Econometrica*, 66, 689–701.

——— (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Arribillaga, R. P., J. Massó, and A. Neme (2016): "Not All Majority-based Social Choice Functions Are Obviously Strategy-proof," .

Arrow, K. J. (1963): *Social Choice and Individual Values*, New York: Wiley, 2nd edition ed.

Ashlagi, I. and Y. A. Gonczarowski (2016): "No stable matching mechanism is obviously strategy-proof." Working Paper.

Bade, S. and Y. Gonczarowski (2016): "Gibbard-Satterthwaite Success Stories and Obvious Strategyproofness," .

[46]Bogomolnaia and Moulin actually prove that Random Priority is the only strategy-proof and ordinally efficient mechanism that satisfies equal treatment of equals when there are three objects and three agents, and with two objects Pareto efficiency and ordinal efficiency become equivalent. The two object case is much simpler than the three object problem that Bogomolnaia and Moulin study, and one can easily verify the above claim without reliance on Bogomlnaia and Moulin's seminal analysis.

Bogomolnaia, A. and H. Moulin (2001): “A New Solution to the Random Assignment Problem,” *Journal of Economic Theory*, 100, 295–328.

Budish, E. and E. Cantillon (2012): “The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard,” *American Economic Review*, 102, 2237–71.

Che, Y.-K. and F. Kojima (2009): “Asymptotic Equivalence of Random Priority and Probabilistic Serial Mechainsms,” Forthcoming, *Econometrica.*

——— (2010): “Asymptotic Equivalence of Random Priority and Probabilistic Serial Mechanisms,” *Econometrica*, 78, 1625–1672.

Dur, U. M. and M. U. Ünver (2015): “Two-Sided Matching via Balanced Exchange: Tuition and Worker Exchanges,” NC State University and Boston College, Working Paper.

Ehlers, L. (2002): “Coalitional Strategy-Proof House Allocation,” *Journal of Economic Theory*, 105, 298–317.

Gibbard, A. (1973): “Manipulation of Voting Schemes: A General Result,” *Econometrica*, 41, 587–601.

——— (1977): “Manipulation of Schemes That Mix Voting with Chance,” *Econometrica*, 45, 665–681.

Hakimov, R. and O. Kesten (2014): “The Equitable Top Trading Cycles Mechanism for School Choice,” WZB and Carnegie Mellon University, Working Paper.

Hatfield, J. W. (2009): “Strategy-Proof, Efficient, and Nonbossy Quota Allocations,” *Social Choice and Welfare*, 33 No. 3, 505–515.

Jones, W. and A. Teytelboym (2016): “Choices, preferences and priorities in a matching system for refugees,” *Forced Migration Review.*

Li, S. (2017): “Obviously Strategy-Proof Mechanisms,” *American Economic Review*, 107, 3257–87.

Liu, Q. and M. Pycia (2013): “Ordinal Efficiency, Fairness, and Incentives in Large Markets,” Working Paper.

Loertscher, S. and L. M. Marx (2015): “Prior-Free Bayesian Optimal Double-Clock Auctions,” Working Paper.

Mackenzie, A. (2017): “A Revelation Principle for Obviously Strategy-proof Implementation,” *working paper.*

Morrill, T. (2014): “Making Just School Assignments,” *Games and Economic Behavior*, (forthcoming).

Pápai, S. (2000): “Strategyproof Assignment by Hierarchical Exchange,” *Econometrica*, 68, 1403–1433.

——— (2001): "Strategyproof and Nonbossy Multiple Assignments," *Journal of Public Economic Theory*, 3, 257–271.

Pycia, M. (2011): "Ordinal Efficiency, Fairness, and Incentives in Large Multi-Unit-Demand Assignments," .

Pycia, M. and M. U. Ünver (2009): "Incentive Compatible Allocation and Exchange of Discrete Resources," UCLA and Boston College, unpublished mimeo.

——— (2016): "Arrovian Efficiency in Allocation of Discrete Resources," UCLA and Boston College, Working Paper.

Rosenthal, R. W. (1981): "Games of perfect information, predatory pricing and the chain-store paradox," *Journal of Economic theory*, 25, 92–100.

Roth, A. (2015): "Migrants Aren't Widgets," *Politico Europe*, http://www.politico.eu/article/migrants–arent–widgets–europe–eu–migrant–refugee–crisis/.

Roth, A. E., T. Sönmez, and M. U. Ünver (2004): "Kidney Exchange," *Quarterly Journal of Economics*, 119, 457–488.

Satterthwaite, M. (1975): "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, 10, 187–216.

Shapley, L. and H. Scarf (1974): "On Cores and Indivisibility," *Journal of Mathematical Economics*, 1, 23–37.

Sönmez, T. and M. U. Ünver (2010): "Course Bidding at Business Schools," *International Economic Review*, 51, 99–123.

Troyan, P. (2016): "Obviously Strategyproof Implementation of Top Trading Cycles," Working paper, University of Virginia.

Zhang, L. and D. Levin (2017): "Partition obvious preference and mechanism design: Theory and experiment," .