Obviously Strategyproof Mechanisms in General Environments

Marek Pycia and Peter Troyan[*]

October 2023

Abstract

We consider the problem of characterizing all obviously strategy-proof (OSP) mechanisms for general preference environments. We show that any OSP mechanism is equivalent to a *generalized millipede game* in which agents are sequentially offered a menu of payoffs they may clinch (and thus leave the game), plus possibility the opportunity to pass (and remain in the game, hoping for better clinching options in the future). Our preference setting unifies many canonical mechanism design settings, such as single-unit auctions, public goods problems, and object allocation, and thus, many of the known OSP mechanisms are special cases of generalized millipede games. We also introduce other examples that fit our preference model that are new to the literature.

1 Introduction

In a mechanism design problem, the goal is to implement a social outcome, where the optimal outcome is a function of the private information of the agents involved. Since the designer does not know this information, they must construct a mechanism—a sequence of actions/announcements taken by the agents, either simultaneously or sequentially—with the final outcome being chosen as a function of the actions taken by all agents. A natural way

[*] Pycia: University of Zurich; Troyan: University of Virginia. The results were first presented in 2019, and reported in the 2020 draft of Pycia and Troyan (2023). Because of length considerations, we split off these results to a separate paper. We thank the audiences at WashU, IFN Stockholm, Ottawa Micro Theory Conference,, the CEME Decentralization Conference in Berkeley, Maryland, Bocconi, Lausanne, Warsaw, Notre Dame, Brown, Glasgow, Michigan, SMU, Essex, SING, UCSD, ISI Delhi, Columbia, Western Ontario, Brown, Richmond Fed, GMU, Israel Online Theory Seminar, LSE Mini-Conference on Behavioural Incentives and Mechanism Design, Collegio Carlo Alberto, and USC. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement No 866376.

to ensure that this outcome is in line with the designer's objectives is for the designer to be confident that they can predict the actions the agents will take for each possible type they may have. In principle, this can be achieved by designing a *strategy-proof* (also known as dominant strategy incentive-compatible) mechanism, which is a mechanism such that each agent i has a single strategy that, no matter what the other agents do, gives i her best possible outcome. However, a recent empirical and experimental evidence suggests that strategyproofness alone may be insufficient if players are unable to recognize that they have a dominant strategy in a given mechanism (either because they have cognitive limitations, or the mechanism is too complex).[1] Thus, they may not play as the designer predicts. Predictability of play requires the mechanism to be sufficiently *simple.*[2]

In a seminal paper, Li (2017) introduces the notion of an *obviously strategy-proof (OSP) mechanism.* He provides a formal characterization of obviously strategy-proof strategies as those that can be recognized as optimal by cognitively limited agents who are unable to engage in the contingent reasoning necessary to recognize dominant strategies in some merely strategyproof mechanisms. He also uses OSP to characterize simple mechanisms for certain binary allocation problems, which include canonical environments such as single-unit auctions. Pycia and Troyan (2023) characterize the full class of OSP mechanisms for a broad class of *no-transfer environments*[3]

In this paper, we consider a wider array of economic environments, and seek a characterization of OSP mechanisms that applies across a broad range of economic settings. The starting point of our model is a set of *outcomes*, $\mathcal{X}$, which can be very general. Each agent i has a preference relation over $\mathcal{X}$ that lies in some domain, $\mathcal{P}_i$. The designer does not know any agent's exact preference, but, depending on the context of the problem, may have some knowledge about how agents view different outcomes. For instance, in a setting with transfers, all else equal, agents prefer an outcome in which they get more money to less; in school choice setting in which students are to be allocated to schools, students care about their own school, but are indifferent to how the other schools are distributed amongst the remaining students. To capture these examples, as well as others, we assume that associated with each $\mathcal{P}_i$ is a binary relation on $\mathcal{X}$, denoted $\trianglerighteq_i$, that we call a *trump relation.*

[1]See, e.g., Kagel, Harstad, and Levin (1987), Hassidim, Romm, and Shorrer (2016), Li (2017), Rees-Jones (2017), Artemov, Che, and He (2017), Shorrer and Sóvágó (2018), and Rees-Jones (2018).

[2]Simplicity has also other benefits including economizing on participation costs (Vickrey, 1961), attracting participants (Spenner and Freeman, 2012), leveling the playing field (Pathak and Sönmez, 2008), and economizing on what the designer needs to know Wilson (1987).

[3]This is the paper on which the current contribution most closely builds. In his seminal work, Li (2017) also provides some results on OSP mechanisms for another canonical mechanism design setting, object assignment problems without transfers. OSP properties of other specific mechanism classes or specific environments include also, e.g., Ashlagi and Gonczarowski (2018), Troyan (2019), Arribillaga et al. (2020), Arribillaga et al. (2019), Bade and Gonczarowski (2017), Mackenzie (2020), Mandal and Roy (2020), Mandal and Roy (2021).

The trump relation is used to model any ex-ante structure the designer knows about agent preferences: if, for two outcomes $x, y \in \mathcal{X}$, we have $x \trianglerighteq_i y$ (read: "x trumps y") then all possible preference types of agent i must (weakly) prefer outcome x to outcome y.[4] Formally, we assume that every preference ranking in the domain is *consistent* with $\trianglerighteq_i$, in the sense that $x \trianglerighteq_i y$ implies $x \succsim_i y$ for all $\succsim_i \in \mathcal{P}_i$. The last key assumption that we make is richness: a preference domain $\mathcal{P}_i$ is *rich* if it contains all strict rankings that are consistent with $\trianglerighteq_i$. We use this model of rich preference domains because it encompasses many disparate economic environments into one unified framework. This will allow us to easily state general results that apply to all rich preference domains, and then immediately derive results for canonical market design applications simply by defining an appropriate trump relation.

Our main contribution is to characterize the full class of OSP mechanisms in all rich preference environments. Specifically, we show that a mechanism is OSP if and only if it is (equivalent to) a *generalized millipede game.* Millipede games were first defined by Pycia and Troyan (2023) for a more restricted class of preference environments. Roughly speaking, a millipede game is an extensive-form game of perfect information in which agents are called to move one at a time. Each time an agent is called to move, she is presented a menu of payoff-equivalent outcomes, or more simply *payoffs*, that she can "clinch". If she selects one of these options, her payoff is determined, and she leaves the game. She may also be given the opportunity to "pass". If she passes, she remains in the game, and may be called again, with a possibly new set of clinching options in the future. To ensure that the game is OSP, when an agent passes, the mechanism must make certain payoff guarantees to her in the future. In the simpler environments of Pycia and Troyan (2023), this amounts to guaranteeing that the next time she moves, she will be given the opporunity to clinch either (i) everything she could have clinched before or (ii) everything that was possible, but not clinchable, at her last move. In this way, the mechanism ensures an agent never "regrets" her decision to pass, in a formal sense as required by OSP.

The main distinction between a millipede game and a generalized millipede game is in the payoff guarantees following a passing action: in a standard millipede game, if a payoff that was previously unclinchable becomes impossible for the agent following a pass, she must be offered all previously clinchable objects. In a generalized millipede game, this need not be the case. The extra structure on the preferences gives the designer extra flexibility in what needs be offered without violating the OSP constraint.

[4]Note that different agents can have different trump relations, and hence, different preference domains.

2 Preliminaries

2.1 Preferences

Let $\mathcal{N} = \{i_1, \ldots, i_N\}$ be a set of agents, and $\mathcal{X}$ a finite set of outcomes. An outcome might involve a monetary transfer; we allow both environments with and without transfers. Each agent has a preference ranking over outcomes, where, for any two $x, y \in \mathcal{X}$, 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 $y \succsim_i x$. For any $\succsim_i$, we let $\succ_i$ denote the corresponding strict preference relation, i.e., $x \succ_i y$ if $x \succsim_i y$ but not $y \succsim_i x$. We use $\mathcal{P}_i$ to denote the domain of agent i's preferences, and will often refer to $\succsim_i$ as agent i's **type**.

2.2 Obvious dominance

Li (2017) introduces the notion of obvious dominance as a solution concept intended to capture what games are simple to play.[5] To start, let Γ be a finite extensive-form game with imperfect information and perfect recall, which is defined in the standard way: there is a finite collection of partially ordered **histories**, $\mathcal{H}$. We write $h' \subseteq h$ to denote that $h' \in \mathcal{H}$ is a subhistory of $h \in \mathcal{H}$, and $h' \subset h$ when $h' \subseteq h$ but $h \neq h'$. Terminal histories will be denoted with bars, i.e., $\bar{h}$. Each $\bar{h} \in \mathcal{H}$ is associated with an outcome in $\mathcal{X}$. At every non-terminal history $h \in \mathcal{H}$, one agent, denoted i_h, is called to play and has a finite set of **actions** $A(h)$ from which to choose. We write $h' = (h, a)$ to denote the history h' that is reached by starting at history h and following the action $a \in A(h)$. To avoid trivialities, we assume that no agent moves twice in a row and that $|A(h)| > 1$ for all non-terminal $h \in \mathcal{H}$. To capture random mechanisms, we also allow for histories h at which a non-strategic agent, Nature, is called to move, and selects an action in $A(h)$ according to some probability distribution.

The set of histories at which agent i moves is denoted $\mathcal{H}_i = \{h \in \mathcal{H} : i_h = i\}$. The set $\mathcal{I}_i$ is a partition of $\mathcal{H}_i$ into **information sets**, where, for any information set $I \in \mathcal{I}_i$ and $h, h' \in I$ 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)

[5] Pycia and Troyan (2023) generalize Li's (2017) definition to what they call simple dominance, which models agents with limited foresight and who thus may only be able to plan for a limited number of moves in the future at any point in a game (rather than constructing a complete contingent plan, as in almost all of the game theory literature, including Li (2017)). While obvious dominance is a special case of simple dominance, the general definition of simple dominance requires additional notation and a novel way to conceptualize what constitutes a "strategy". In this paper, we are concerned only with characterizing OSP mechanisms, and so we present just the simpler definition of obvious dominance, which only requires the standard game-theoretic notion of a strategy as a complete contingent plan of action.

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 use $A(I)$ to denote the set of actions available at information set I, and denote by $I(h) \in \mathcal{I}_i$ the information set containing history h. We say that an information set I_1 **precedes** information set I_2 if there are $h_1 \in I_1$ and $h_2 \in I_2$ such that $h_1 \subseteq h_2$; we then write $I_1 \leq I_2$ (and $I_1 < I_2$ if $I_1 \neq I_2$) and we also say that I_2 **follows** I_1 and that I_2 is a **continuation** of I_1. We say that an outcome x is **possible** at information set I if there is $h \in I$ and a terminal history $\bar{h} \supseteq h$ such that x obtains at $\bar{h}$.

A **strategy** for an agent in Γ, S_i, is a function that maps each information set $I \in \mathcal{I}_i$ to an action in $A(I)$.[6] We write $S_i(\succ_i)$ when we want to refer to the strategy of a particular type $\succ_i$ of agent i, and $S_i(\succ_i)(I)$ for the action taken at information set I when agent i under strategy $S_i(\succ_i)$. We use $S_{\mathcal{N}}(\succ_{\mathcal{N}}) = (S_i(\succ_i))_{i \in \mathcal{N}, \succ_i \in \mathcal{P}_i}$ to denote a profile of (type) strategies. A **mechanism** $(\Gamma, S_{\mathcal{N}})$ is an extensive-form game together with a profile of strategies. Two mechanisms $(\Gamma, S_{\mathcal{N}})$ and $(\Gamma', S'_{\mathcal{N}})$ are said to be **equivalent** if, for every profile of types $\succ_{\mathcal{N}}$, the distribution of outcomes when the agents follow $S_{\mathcal{N}}(\succ_{\mathcal{N}})$ in Γ is the same as when they follow $S'_{\mathcal{N}}(\succ_{\mathcal{N}})$ in Γ'. Note that this equivalence definition is purely outcome-based. Given a mechanism, we can construct the corresponding **social choice rule**—that is, mapping from preference profiles to outcomes—that is implemented. All mechanisms in the same equivalence class implement the same social choice rule.

We are now ready to define the main solution concept of obvious dominance. A strategy $S_i(\succ_i)$ is **obviously dominant** for type $\succ_i$ of agent i if, at all $I^* \in \mathcal{I}_i$, in the continuation game starting at I^*, the worst possible outcome when i follows $S_i(\succ_i)$ at all $I \in \mathcal{I}_i$ is weakly preferred by type $\succ_i$ to the best possible outcome in the continuation game when i follows some other action $a' \neq S_i(\succ_i)(I^*)$ at I^*. If a game Γ admits a profile of strategies $S_{\mathcal{N}}(\succ_{\mathcal{N}})$ such that $S_i(\succ_i)$ is obviously dominant for all types $\succ_i \in \mathcal{P}_i$ and all agents i, then we say that the mechanism $(\Gamma, S_{\mathcal{N}})$ is **obviously strategy-proof (OSP)**.

3 Preference Environments

The preference model specified in the previous section is very general. At the same time, in many practical market design applications, more is known about the structure of agents' preferences, which determines their preference domains $\mathcal{P}_i$. For instance, in an auction environment, it is known that agents prefer more money to less, and this should be taken into account when defining their preference domains. Our goal is to cover as many different applications as possible in one simple, unified framework, so that results for specific applications will follow as special cases of our general theorems. In this section, we explain how this is

[6]We focus on pure strategies; the extension to mixed strategies is straightforward.

done using what we call a *trump relation* to define the set of admissible preference profiles in the domain $\mathcal{P}_i$.

3.1 Structural preferences and richness

For each agent i, there is an associated reflexive and transitive binary relation on the set $\mathcal{X}$, denoted $\trianglerighteq_i$; we call $\trianglerighteq_i$ a **trump relation**. The motivation is that the designer may have some information about the structure of an agent's preference, and knows that some outcomes are better than (trump) others, for all possible types of agent i. For instance, in environments with transfers, outcome x trumps outcome y ($x \trianglerighteq_i y$) for agent i if i receives a higher transfer under x than y, and all else is equal. When $x \trianglerighteq_i y$ but not $y \trianglerighteq_i x$, we say that x **strictly trumps** y. Varying the trump relations $\trianglerighteq_i$ is what will allow us to easily capture many different preference settings as special cases of our result (concrete examples will be given below).

Trump relations are primitives of the model that capture the structure of possible agent preferences; that is, the designer knows each agent's trump relation, and hence her preference domain $\mathcal{P}_i$, but not her precise preference $\succsim_i$ in the domain, which may be any preference that is *consistent* with $\trianglerighteq_i$. Formally, a preference ranking $\succsim_i \in \mathcal{P}_i$ is **consistent** with $\trianglerighteq_i$ if, for any $x, y \in \mathcal{X}$, $x \trianglerighteq_i y$ implies $x \succsim_i y$, and $x \triangleright_i y$ implies $x \succ_i y$.[7] Given some $\trianglerighteq_i$, we assume that all admissible preference rankings $\succsim_i \in \mathcal{P}_i$ are consistent with $\trianglerighteq_i$. The trump relations for each agent $\trianglerighteq_i$ are taken as a primitive of the model, and we allow the possibility that different agents have different relations, and therefore different preference domains.

If $x \trianglerighteq_i y$ and $y \trianglerighteq_i x$ then x and y are $\trianglerighteq_i$ −**equivalent**. Any $\trianglerighteq_i$ determines an **equivalence partition** of $\mathcal{X}$. We refer to each element of the equivalence partition as a **payoff** of the agent in question. When the distinction between a payoff and an outcome is important we write $[x]_i = \{y \in \mathcal{X} : x \trianglerighteq_i y \text{ and } y \trianglerighteq_i x\}$ to represent the payoff (the element of the partition) that contains x. Elsewhere, to avoid unnecessary formalism, we will write "payoff x", or simply just "x", to refer to the partition element to which outcome x belongs; thus phrases such as "payoff x obtains" are understood as "some $y \in [x]_i$ obtains". With slight abuse of notation, we will extend $\trianglerighteq_i$ and $\succsim_i$ to be defined over payoffs in the natural way, and write, e.g., $x \succ_i y$ to denote that payoff x is strictly prefer to payoff y. A payoff x is said to be **untrumped** in a subset of payoffs for agent i if there is no payoff y in this subset such that $y \triangleright_i x$.

The key assumption we make on preference domains is richness, where we say that $\mathcal{P}_i$ is **rich** if it contains all strict rankings over payoffs that are consistent with $\trianglerighteq_i$. We are

[7]When the trump relation takes the form $x \trianglerighteq_i y$ if and only if $x = y$, any domain is consistent with $\trianglerighteq_i$.

agnostic as to whether consistent non-strict rankings belong to $\mathcal{P}_i$ or not.[8] We make this assumption because it allows us to easily state general results that apply to all rich preference domains, and then immediately derive results for canonical market design applications simply by defining an appropriate trump relation. The richness assumption is very flexible, and encompasses many standard economic environments, including both those with and without transfers. Examples of no-transfer environments include voting (all agents have strict preferences over $\mathcal{X}$) and the allocation of indivisible goods (e.g., school choice). An example of environments with transfers that are captured is auctions of either a single or multiple goods.[9]

Given the definition of obvious strategy-proofness, it is natural to ask precisely which mechanisms are OSP. Pycia and Troyan (2023) characterize the class of obviously dominant mechanisms for the special case when the trump relations $\trianglerighteq_i$ are symmetric for all i.[10] This captures many interesting problems. One example is the classical voting environment (e.g., Gibbard (1973) and Satterthwaite (1975)), in which agents have strict rankings over all outcomes in $\mathcal{X}$. Another is the allocation of indivisible objects without transfers, in which each $x \in \mathcal{X}$ represents the complete allocation of all of the objects to the agents, but agents only care about their individual allocation, and are indifferent between how the remaining goods are allocated across the other agents.[11] One prominent application that fits into this latter setting is the *school choice problem* (Abdulkadiroğlu and Sönmez, 2003), where the agents are students and the objects are schools.

While many environments do satisfy the symmetry requirement, there are also economically interesting settings where it is violated. For instance, symmetry rules out all settings with transfers. This is because if $x, y \in \mathcal{X}$ are two outcomes that are equivalent in all respects except that some agent i receives a higher transfer under x than y, then x trumps y ($x \trianglerighteq_i y$) but y does not trump x ($y \ntrianglerighteq_i x$), and so $\trianglerighteq_i$ is not symmetric. Therefore, the characterization theorem of Pycia and Troyan (2023) does not apply to any setting with transfers.[12]

[8] Our use of the term richness shares with other uses of the term in the literature the idea that the domain of preferences contains sufficiently many profiles: if certain preference profiles belong to the domain, then some other profiles belong to it as well (cf. Dasgupta, Hammond, and Maskin (1979) and Pycia (2012)).

[9] For details on how to define the relevant trump relations for each of these environments, see Pycia and Troyan (2023). While richness is very flexible, not all preference domains are rich. For instance, domains of single-peaked preferences are typically not rich, and Arribillaga, Massó, and Neme (2020) show that some of our results do not extend to single-peaked preference domains.

[10] Recall that a binary relation $\trianglerighteq_i$ is **symmetric** if $x \trianglerighteq_i y$ implies $y \trianglerighteq_i x$.

[11] The voting environment is captured by defining $\trianglerighteq_i$ such that $x \trianglerighteq_i y$ if and only if $x = y$ for all i, and richness implies that every strict ranking over x belongs to $\mathcal{P}_i$. Object allocation is captured by defining $\trianglerighteq_i$ such that $x \trianglerighteq_i y$ if and only if agent i receives the same set of goods in allocations x and y. In this case, each element of agent i's equivalence partition of $\mathcal{X}$ can be identified with the set of objects she receives, and richness implies that every strict ranking of the sets of possible objects belongs to $\mathcal{P}_i$.

[12] Li (2017) characterizes OSP mechanisms for the specific environment of binary allocation problems with

Further, it is not just transfers that are ruled out by symmetry. In a standard object allocation problem without transfers such as school choice, the trump relations are symmetric, and richness implies that agents may have any strict over the objects. However, sometimes the designer may have additional information about the student/agent preferences. For instance, it may be known that schools can be divided into different quality tiers, or that preferences depend on geography, as in the following example.

Example 1. There are 6 students $\mathcal{N} = \{i_1, i_2, i_3, i_4, i_5, i_6\}$ and 6 schools, $\mathcal{S} = \{H, M, P, S, B, L\}$, which, for concreteness, may be thought of as Harvard, MIT, Princeton, Stanford, Berkeley, and UCLA, respectively. Each student can be either an "East Coast student", with preference domain $\mathcal{P}_E$, or a "West Coast student", with preference domain $\mathcal{P}_W$. The East Coast students prefer to go to any school on the East Coast (H, M, P) to any school on the West Coast (S, B, L), but may have any preference among the set of East Coast schools (and any preference among the set of West Coast schools). Formally, the preference domain $\mathcal{P}_E$ is governed by the trump relation $\trianglerighteq_E$ defined such that $s \trianglerighteq_E s'$ if and only if

1. $s = s'$, or
2. $s \in \{Harvard, MIT, Princeton\}$ and $s' \in \{Stanford, Berkeley, UCLA\}$;

otherwise, $s \ntrianglerighteq_E s'$. Given this trump relation, the domain $\mathcal{P}_E$ satisfies:

1. Consistency: all $\succsim \in \mathcal{P}_E$ are consistent with $\trianglerighteq_E$;[13]
2. Richness: $\mathcal{P}_E$ contains all strict rankings of the schools that are consistent with $\trianglerighteq_E$.

The domain $\mathcal{P}_W$ for West Coast students is defined analogously, switching the roles of the East Coast and West Coast schools.

It is easy to see that this is not a symmetric environment. For example, consider an East Coast student i such that $\trianglerighteq_i = \trianglerighteq_E$ and so $\mathcal{P}_i = \mathcal{P}_E$. Let $s = Harvard$ and $s' = Stanford$. Then, $s \trianglerighteq_E s'$, but $s' \ntrianglerighteq_E s$, and thus $\trianglerighteq_E$ is not a symmetric relation. Thus, the known characterizations of OSP mechanisms will not apply to this environment, and a more general result is needed.

transfers as the class of personal clock auctions. As we show later in the paper, binary allocation problems with transfers can be modeled in our framework using a particular choice of $\trianglerighteq_i$, and personal clock auctions are a special case of generalized millipede games that are our main object of study.

[13]For example, preference rankings such as $P \succ_i M \succ_i H \succ_i B \succ_i L \succ_i S$ and $M \succ_i P \succ_i H \succ_i B \succ_i L \succ_i S$ belong to $\mathcal{P}_E$, but preference rankings such as $P \succ_i B \succ_i H \succ_i M \succ_i L \succ_i S$ do not.

4 Millipede Games

Pycia and Troyan (2023) introduce a class of games called *millipede games*.[14] Roughly speaking, a millipede game is an extensive-form game of perfect information that satisfies the following properties, which can be thought of as "structural properties" and "incentive properties":

1. **Structure:** The set of actions can be divided into *clinching actions* and *passing actions*. If an agent selects a clinching action, she never moves again in the game, and her payoff is completely determined. If an agent selects a passing action, she may be called to move again, and multiple payoffs are still possible for her. There can be at most one passing action at each history.

2. **Incentives:** Following a passing action, the agent who passes is given certain *payoff guarantees* that govern the set of payoffs she may receive in the future of the game.

The requirements on the payoff guarantees are such that it is optimal (in a precise sense) for an agent to choose the passing action if her top still-possible payoff is not clinchable. The exact nature of these guarantees are determined by the preference environment and simplicity concepts imposed by the designer. For instance, Pycia and Troyan (2023) provide the necessary condition on payoff guarantees to ensure the resulting class of millipede games are all obviously dominant in symmetric preference environments. The main result of this section gives a more general condition that applies to all rich preference environments.

To formally define a millipede game, first say that payoff x is **possible** for agent i at history h if there is a terminal history $\bar{h} \supseteq h$ such that at the outcome associated with $\bar{h}$, agent i obtains payoff x. For any history h, $P_i(h)$ denotes the set of payoffs that are possible for i at h. We say that agent i has **clinched** payoff x at history h if agent i is not called to move at any $h' \supseteq h$, and at all terminal histories $\bar{h} \supseteq h$, agent i receives payoff x. For $h \in \mathcal{H}_i$, we denote by $C_i(h)$ the set of all payoffs x that i **can clinch** (or that are **clinchable**) at h; that is, $C_i(h)$ is the set of payoffs for which there is an action $a \in A(h)$—referred to as a **clinching action**—such that i has clinched x at the history (h,a). At a terminal history $\bar{h}$, no agent is called to move and there are no actions; however, it will be notationally useful to define $C_i(\bar{h}) = \{x\}$, where x is the payoff that i obtains at terminal history $\bar{h}$.

We further define $C_i^{\subseteq}(h) = \{x : x \in C_i(h')$ for some $h' \subseteq h$ s.t. $i_{h'} = i\}$ to be the set of payoffs that i can clinch at some subhistory of h, and $C_i^{\subsetneq}(h) = \{x : x \in C_i(h')$ for some $h' \subsetneq$

[14]Ashlagi and Gonczarowski (2018), Troyan (2019), Bade and Gonczarowski (2017), and Mandal and Roy (2020) provide characterizations of OSP mechanisms for the specific problem of indivisible object allocation without transfers. These are all special cases of millipede games.

h s.t. $i_{h'} = i\}$ to be the set of payoffs that i can clinch at some *strict* subhistory of h. Note that while the definition of $C_i(h)$ presumes that i moves at h or h is terminal, the payoff sets $P_i(h)$, $C_i^{\subseteq}(h)$ and $C_i^{\subset}(h)$ are well-defined for any h, whether i moves at h or not, and whether h is terminal or not.

Finally, consider a history h such that $i_{h'} = i$ for some $h' \subsetneq h$ and either $i_h = i$ or h is a terminal history. We say that payoff x **becomes impossible** for i at h if (i) for all $h' \subsetneq h$, there exists some $x' \trianglerighteq_i x$ such that $x' \in P_i(h')$ and (ii) for all $x' \trianglerighteq_i x$, $x' \notin P_i(h)$. A payoff x is **previously unclinchable** at h if, for all $x' \trianglerighteq_i x$, $x' \notin C_i^{\subset}(h)$.[15]

Definition 1. A **millipede game** is a finite extensive-form game of perfect information that satisfies the following properties:

1. Nature either moves once, at the empty history $h_{\varnothing}$, or Nature has no moves.

2. At any history at which an agent moves, all but at most one action are clinching actions, and the remaining action–if there is one–is a passing action (there may be several clinching actions associated with the same payoff).

3. At all h, if there exists a previously unclinchable payoff z that becomes impossible for agent i_h at h, then for all $x \in C_{i_h}^{\subset}(h)$: (i) if h is terminal, then $y \trianglerighteq_i x$, where y is the payoff for i that obtains at h, or (ii) if h is a history where agent i_h moves, then there exists an action $a \in A(h)$ such that for all $y \in P_{i_h}((h, a))$, we have $y \trianglerighteq_i x$.

The first two parts of the definition are conditions on the structure of the game tree. The third part formalizes the idea of "payoff guarantees" following a passing move. In symmetric preference environments, all payoffs are untrumped, and Part 3 reduces to the condition of Pycia and Troyan (2023) that if a previously unclinchable payoff x becomes impossible, then the agent must be offered the opportunity to clinch all payoffs she could have clinched previously. An equivalent way to re-state this is that if an agent passes, the next time she moves, she must be given the opportunity to clinch either (i) everything that she could have clinched before or (ii) everything that was possible, but not clinchable, at her last move.

In more general (i.e., non-symmetric) preference environments, part 3 says that if payoff z becomes impossible at h, then for every x that i could have clinched previously, she must be able to guarantee either x or some y that trumps x. There are two subtleties here that are worth emphasizing. The first is that is that does not say that i must be able to clinch y

[15] The definitions of clinchable/possible payoffs refer to a single payoff x, while the definition of becoming impossible refers to the class of all payoffs that trump x. When $\trianglerighteq_i$ is symmetric, this is a singleton, and so x becomes impossible when x itself disappears. In more general environments, there may be a larger class of payoffs that trump x, and so x becomes impossible only when all payoffs x and everything that trumps x are no longer possible. The same comment applies to a payoff being previously unclinchable.

or something better, but only that she must be able to guarantee that she will get something better than y. For instance, if i's trump relation is such that $y'' \rhd_i y' \rhd_i y$, then it known that all types of agent i must be such that $y'' \succ_i y' \succ_i y$. Then, it is feasible that there is an action at h that does not clinch either y'' or y', but following this action, i will get either y'' or y'.

Our characterization of OSP games and mechanisms involves the simple following strategy: given a type $\succ_i$, we call strategy $S_i(\succ_i)$ a **greedy strategy** if, for any h at which i can clinch her best still-possible payoff x, the action $S_i(\succ_i)(h)$ is one that clinches x for the agent; otherwise, the agent passes. We refer to millipede games with greedy strategies as millipede mechanisms.

Theorem 1. *Every OSP mechanism is equivalent to a generalized millipede mechanism. Every generalized millipede mechanism is OSP.*

Since millipede mechanisms are a special case of generalized millipede mechanisms, it is immediate that anything that can be implemented by a millipede mechanism can be implemented by a generalized millipede mechanism. In some environments in which at least one agent's trumping relation is not symmetric, strictly more social choice rules can be implemented via generalized millipede mechanisms relative to what can be implemented by millipede mechanisms. The examples presented in the next section illustrate some aspects of the increased flexibility of generalized millipede mechanisms.

Examples of generalized millipede mechanisms

We now present three examples of generalized millipede games in specific environments.

Example 2. The first example we consider is the allocation of a single good with transfers. For this problem, one of the most studied mechanisms is the ascending auction, in which there is a going price, and at each step, agents are asked whether they want to continue in the auction at that price, or quit. The price rises until all agents but one have quit, and the remaining agent gets the object at the final price.

Li (2017) shows that ascending auctions are OSP; indeed, ascending auctions are also a special case of generalized millipede mechanisms. Part 1 of Definition 1 is trivial. For Part 2, note that at each step, an agent has one passing action ("continue") and one clinching action ("quit", or "clinch a payoff of zero"), and so Part 2 is also satisfied. Part 3 follows from the fact that every time the price rises, an agent is offered the opportunity to quit.

A generalization of the single-unit auction is what Li (2017) refers to as binary allocation problems, defined as follows. The set of outcomes is $\mathcal{X} = Y \times \mathbb{R}^{\mathcal{N}}$, where $Y \subseteq \{0,1\}^{\mathcal{N}}$ is a set of feasible allocations and $\mathbb{R}^{\mathcal{N}}$ is the set of profiles of transfers, one for each agent; a generic allocation is denoted y and a generic profile of transfers $w = (w_i)_{i \in \mathcal{N}}$. In this section, we denote types by $\theta_i \in [\underline{\theta}_i, \overline{\theta}_i]$, where $0 \leq \underline{\theta}_i < \overline{\theta}_i < \infty$, and assume each agent has preferences represented by a quasilinear utility function: $u_i(\theta_i, y, w) = \theta_i y_i + w_i$.[16]

This framework captures many important environments of economic interest, including single-unit auctions, procurement auctions, and binary public goods games. Li (2017) shows that in these environments, every OSP mechanism is equivalent to a *personal clock auction*. A personal clock auction is similar to an ascending auction, but generalizes it in the following ways: first, agents may face different individualized prices ("clocks"); second, at any point, there may be multiple quitting actions that allow agents to drop out of the auction, or multiple continuing actions that allow them to stay in the auction; and third, when an agent quits, her transfer need not be zero. The key restrictions are that each agent's clock must be monotonic, and that whenever the personal price an agent faces strictly changes, she must be offered an opportunity to quit.

It is straightforward to verify that this environment satisfies richness, with the trump relation $\trianglerighteq_i$ for agent i defined as follows: $(y, w) \trianglerighteq_i (y', w')$ if and only if $w_i \geq w'_i$ and $y_i \geq y'_i$. Therefore, Theorem 1 applies. Indeed, personal clock auctions are a special case of generalized millipede games (for essentially the same reasons as outlined for ascending auctions above), and so we recover Li's (2017) result as a special case of our general theorem.

The next example highlights that the added generality of Theorem 1 goes beyond just adding transfers.

Example 3. Consider the preference environment of Example 1, in which i_1 is a West Coast student with preference domain $\mathcal{P}_W$ and i_2 is an East Coast student with preference domain $\mathcal{P}_E$.[17] Figure 1 shows an example of a generalized millipede game for this environment. For brevity, we present only the beginning of the game that assigns students i_1 and i_2, and use dots to indicate the continuation game on the remaining students. We present this example because it is a nontrivial instance of a generalized millipede game that is not a millipede game in the original sense defined in Pycia and Troyan (2023). To see this, consider the history following i_1 passing and i_2 clinching H, where i_1 is now offered a choice between S and B. At this history, L is a previously unclinchable payoff that has become impossible, so, a standard millipede game would require i_1 to be offered H, M and P. But, H has already

[16] Note that this assumes a continuum of types and transfers, which we do in order to reproduce the binary allocation environment of Li (2017). The definition of generalized millipede games extends trivially to this set up.

[17] The game in Figure 1 remains OSP when the planner does not know which are the East Coast and which are the West Coast students, and desiring to design a game that is OSP no matter which type (East or West) each student turns out to be.

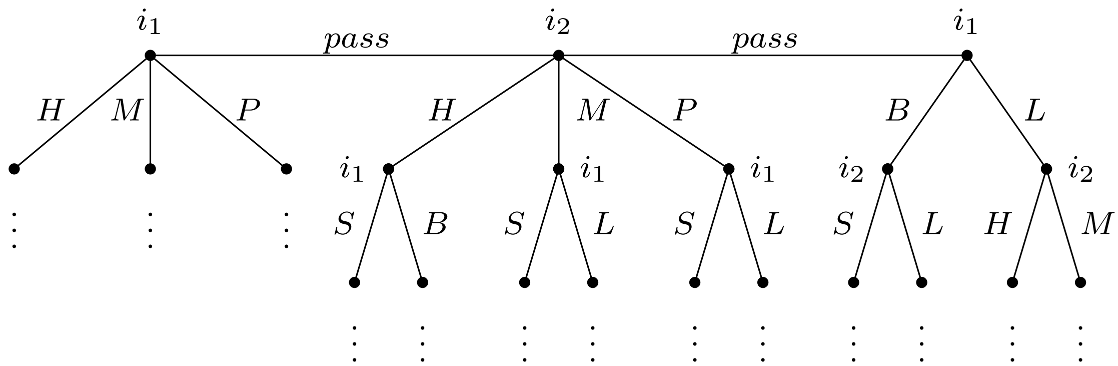

Figure 1: A generalized millipede game for the preference environment in Example 1. The dots indicate that the game continues in some manner consistent with a generalized millipede game (for instance, one possibility is a serial dictatorship on the remaining unassigned agents).

been claimed by i_2 here, and so this is not possible. It is the extra preference structure that allows us to offer only S and B, while still retaining OSP: if i_1 passes on H, M and P initially, then the mechanism is able to infer that she is a West Coast student, and so $S, B \succ_{i_1} H, M, P$. Therefore, her worst case from passing at the initial node is some West Coast school, which is better than any East Coast school, and so the game is OSP.

In the previous example, the agents' preference domains were asymmetric: some were East Coast students, and some were West Coast students. As a final example, we show that this feature is not necessary for constructing nontrivial generalized millipede games. This also highlights the generality of our model by giving another example of an instance that fits into the preference setting.

Example 4. There are a set of public schools $\mathcal{S} = \{H_1, H_2, L_1, L_2, A\}$, and 5 students, $\mathcal{N} = \{i_1, i_2, i_3, i_4, i_5\}$. All students have the same preference domain $\mathcal{P}$, governed by a trump relation $\trianglerighteq$ constructed as follows: $s \trianglerighteq s'$ if and only if either (i) $s = s'$ or (ii) $s \in \{H_1, H_2\}$ and $s' \in \{L_1, L_2\}$. The interpretation is that the H schools are "high-performing" while the L schools are "low-performing", and all students prefer a high-performing school to a low-performing school. The remaining school, A, is not $\trianglerighteq$ −related to any other school. We can think of A as a special program, such as an Arts school, which can rank anywhere in an agent's preferences: if a student is an Arts student, then she may rank school A highly, but if not, she may rank school A low, and it is not known to the designer which students are the Arts students and which are not.

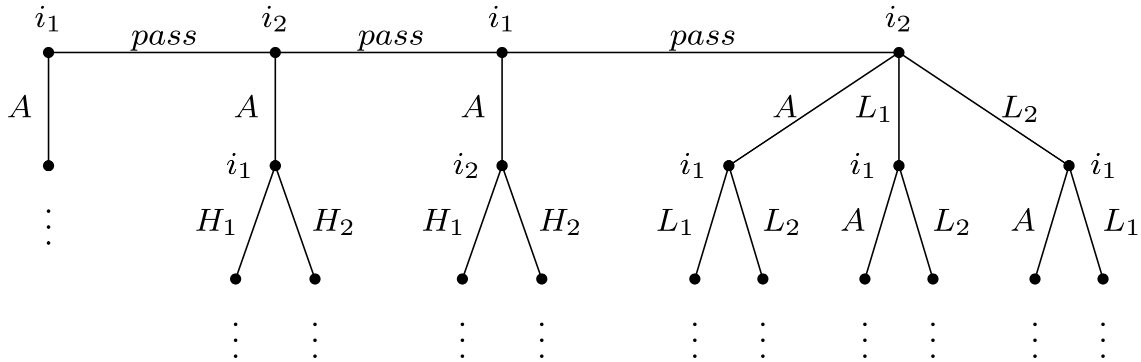

Figure 2: A generalized millipede game for Example 2. The dots indicate that the game continues in some manner consistent with a generalized millipede game (for instance, a serial dictatorship on the remaining unassigned agents).

Figure 2 shows a generalized millipede game for this environment. Once again, this game is not a standard millipede game in the sense of Pycia and Troyan (2023). To see this, note that at the history where i_1 is offered a choice between H_1 and H_2, payoffs L_1 and L_2 both become impossible, and were both previously unclinchable. Thus, a (non-generalized) millipede game would require that i_1 be offered everything that was previously clinchable. This means that i_1 would need to be offered A, which cannot happen here because A has already been claimed by i_2. In a symmetric environment (e.g., object allocation without the additional structure imposed here), this would violate OSP. In this environment, though, when i_1 passes on A, because of the structural preference $\trianglerighteq$, the mechanism can infer that she is not an Arts student, and so i_1's top choice must be either H_1 or H_2. Thus, this is all that needs to be offered to ensure the worst-case criterion is satisfied. The more expansive definition of a generalized millipede game presented in this paper expands the class in precisely a way that allows for this possibility.

Note also that, while there are no transfers, this game does have the flavor of an ascending auction: each time an agent is called to play, she is offered A, which is akin to being offered the opportunity to quit in an ascending auction. In an auction, this price discovery allows the mechanism to infer the agent's value for the object. Here, the preference structure is such that the relative rankings of $\{H_1, H_2\}$ and $\{L_1, L_2\}$ are known, but nothing is known about A. As the mechanism progresses, it learns more about the "value" of A relative to the other objects: If i_1, say, clinches A the first time it is offered, it must be her top-ranked object; if i_1 passes, then the mechanism now infer that A must be ranked below $\{H_1, H_2\}$ (because if i_1 passes, the next time she moves, it may be that both H_1 and H_2 have disappeared as a possibility for her). As the agents continue to pass on A, it becomes known that A moves further down in their preference lists.

References

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

Arribillaga, R. P., J. Massó, and A. Neme (2019): "All Sequential Allotment Rules Are Obviously Strategy-Proof," BGSE Working Paper 1108.

——— (2020): "On obvious strategy-proofness and single-peakedness," *Journal of Economic Theory*, 186, 1–34.

Artemov, G., Y.-K. Che, and Y. He (2017): "Strategic 'Mistakes': Implications for Market Design Research," Tech. rep., mimeo.

Ashlagi, I. and Y. A. Gonczarowski (2018): "Stable matching mechanisms are not obviously strategy-proof," *Journal of Economic Theory*, 177, 405–425.

Bade, S. and Y. Gonczarowski (2017): "Gibbard-Satterthwaite Success Stories and Obvious Strategyproofness," .

Dasgupta, P., P. Hammond, and E. Maskin (1979): "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, 46, 185–216.

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

Hassidim, A., A. Romm, and R. I. Shorrer (2016): "Strategic behavior in a strategy-proof environment," in *Proceedings of the 2016 ACM Conference on Economics and Computation*, ACM, 763–764.

Kagel, J. H., R. M. Harstad, and D. Levin (1987): "Information impact and allocation rules in auctions with affiliated private values: A laboratory study," *Econometrica: Journal of the Econometric Society*, 1275–1304.

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

Mackenzie, A. (2020): "A Revelation Principle for Obviously Strategy-proof Implementation," *Games and Economic Behavior*, 124, 512–533.

Mandal, P. and S. Roy (2020): "Obviously Strategy-proof Implementation of Assignment Rules: A New Characterization," .

——— (2021): "On Obviously Strategy-proof Implementation of Fixed Priority Top Trading Cycles with Outside Options," *Economics Letters*, Forthcoming.

Pathak, P. A. and T. Sönmez (2008): "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *American Economic Review*, 98, 1636–1652.

Pycia, M. (2012): "Stability and Preference Alignment in Matching and Coalition Formation," *Econometrica*, 80, 323–362.

Pycia, M. and P. Troyan (2023): "A Theory of Simplicity in Games and Mechanism Design," *Econometrica*, forthcoming.

Rees-Jones, A. (2017): "Mistaken play in the deferred acceptance algorithm: Implications for positive assortative matching," *American Economic Review*, 107, 225–29.

——— (2018): "Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match," *Games and Economic Behavior*, 108, 317–330.

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.

Shorrer, R. I. and S. Sóvágó (2018): "Obvious mistakes in a strategically simple college admissions environment: Causes and consequences," *Available at SSRN 2993538*.

Spenner, P. and K. Freeman (2012): "To Keep Your Customers, Keep It Simple," *Harvard Business Review*, May.

Troyan, P. (2019): "Obviously Strategy-Proof Implemenation of Top Trading Cycles," *International Economic Review*, 60.

Vickrey, W. (1961): "Counterspeculation, Auctions and Competitive Sealed Tenders," *Journal of Finance*, 16, 8–37.

Wilson, R. (1987): "Game-Theoretic Approaches to Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley, Cambridge: Cambridge University Press, 33–77.

A Proof of Theorem 1

The proof of this theorem builds on the proof of Theorem 5 of Pycia and Troyan (2023). While similar in structure, the proof in Pycia and Troyan (2023) is only valid for symmetric trump relations, and additional arguments are needed to account for more general preference environments. Recall that $P_i(h)$ and $C_i(h)$ denote the sets of possible and clinchable payoffs, respectively, for agent i at history h. We further say payoff x is **guaranteeable** for i at h if there is some continuation strategy S_i such that i receives payoff x at all terminal histories $\bar{h} \supseteq h$ that are consistent with i following S_i. We use $G_i(h)$ to denote the set of payoffs that are guaranteeable for i at history h.

We also make use of the Li's (2017) Pruning Principle, which says that, starting with any OSP mechanism (Γ, S), we can create a new game Γ' by deleting all histories of Γ that are never reached for any type profile under S—called the **pruned game**—such that the restriction of S to Γ' is obviously dominant for Γ', and both games result in the same outcome. (See Li (2017) for a more formal statement.)

We break this proof up into two propositions, corresponding to the two statements in the theorem.

Proposition 1. *Every OSP mechanism is equivalent to a generalized millipede mechanism with greedy strategies.*

Proof of Proposition 1. To begin, note that it is without loss of generality to consider only perfect information OSP games in which Nature moves at most once, as the first mover;[18] this corresponds to part 1 in the definition of a generalized millipede game (Definition 1). We now prove two lemmas that correspond to parts 2 and 3 of Definition 1.

Lemma 1. *For any OSP game Γ, there exists an equivalent OSP game Γ' such that: (i) at each h, at least $|A(h)| - 1$ actions are clinching actions; (ii) for every payoff $x \in G_i(h)$, there exists an action $a_x \in A(h)$ that clinches x for i; and, (iii) if $P_i(h) = G_i(h)$, then all actions in $A(h)$ are clinching actions, and $i_{h'} \neq i$ for any $h' \supsetneq h$.*

Proof. We begin by proving the following claim, which says that there is at most one action such that, following this action, not all possible payoffs are guaranteeable.[19]

[18] This simple observation was first pointed out in a footnote by Ashlagi and Gonczarowski (2018); Pycia and Troyan (2023) provide the complete proof. See also **?**, who generalizes this even further by relaxing recall requirements.

[19] A related claim appears as Lemma A2 of Pycia and Troyan (2023). Their proof is not valid for non-symmetric trump relations.

Claim 1. Let Γ be an obviously strategy-proof game of perfect information that is pruned with respect to the obviously dominant strategy profile $S_{\mathcal{N}}$. Consider a history h where agent $i_h = i$ is called to move. There is at most one action $a^* \in A(h)$ such that $P_i((h, a^*)) \notin G_i(h)$.

Proof of Claim 1. For any history h, let $PnG_i(h) = P_i(h) \setminus G_i(h)$ (where "**PnG**" is shorthand for "possible but not guaranteeable"). Now, consider any h at which i moves, and assume that at h, there are (at least) two such actions $a_1^*, a_2^* \in A(h)$ as in the statement. We first claim that $PnG_i(h) \cap P_i(h_1^*) \cap P_i(h_2^*) = \varnothing$, where $h_1^* = (h, a_1^*)$ and $h_2^* = (h, a_2^*)$. Indeed, if not, then let x be a payoff in this intersection. By pruning, some type $\succ_i$ is following some strategy such that $S_i(\succ_i)(h) = a_1^*$ that results in a payoff of x at some terminal history $\bar{h} \supseteq (h, a_1^*)$. Note that $Top(\succ_i, P_i(h)) \neq x$, because otherwise a_1^* would not be obviously dominant for this type (since $x \notin G_i(h)$ and $x \in P_i(h_2^*)$). Thus, let $Top(\succ_i, P_i(h)) = y$. Note that $y \notin G_i(h)$ (or else it would not be obviously dominant for type $\succ_i$ to play a strategy such that x is a possible payoff). Further, we must have $y \in P_i(h_1^*)$ and $y \notin P_i(h_2^*)$. To see the former, note that if $y \notin P_i(h_1^*)$, then a_1^* is not obviously dominant for type $\succ_i$, which contradicts that $S_i(\succ_i)(h) = a_1^*$; given the former, if $y \in P_i(h_2^*)$, then once again a_1^* would not be obviously dominant for type $\succ_i$. Now, again by pruning, there must be some type $\succ_i'$ such that $S_i(\succ_i')(h) = a_2^*$ that results in payoff x at some terminal history $\bar{h} \supseteq (h, a_2^*)$. By similar reasoning as previously, $Top(\succ_i', P_i(h)) \neq x$, and so $Top(\succ_i', P_i(h)) = z$ for some $z \in P_i(h_2^*)$. Since $y \notin P_i(h_2^*)$, we have $z \neq y$, and we can as above conclude that $z \notin G_i(h)$. It is without loss of generality to consider a type $\succ_i'$ such that $Top(\succ_i', P_i(h) \setminus \{z\}) = y$.[20] Note that, for this type, no action $a \neq a_2^*$ can obviously dominate a_2^* (since $z \notin G_i(h)$). Further, a_2^* itself is not obviously dominant for this type, since the worst case from a_2^* is strictly worse than y (since $y \notin P_i(h_2^*)$ and $z \notin G_i(h)$), while $y \in P_i(h_1^*)$. Therefore, this type has no obviously dominant action at h, which is a contradiction.

Thus, $PnG_i(h) \cap P_i(h_1^*) \cap P_i(h_2^*) = \varnothing$, which means there must be distinct x, y such that (i) $x, y \in PnG_i(h)$ (ii) $x \in P_i(h_1^*)$ but $x \notin P_i(h_2^*)$ and (iii) $y \in P_i(h_2^*)$ but $y \notin P_i(h_1^*)$. Also, note that $y' \rhd_i y$ implies $y' \notin P_i(h_1^*)$—indeed, if there were such a y', then it would not be obviously dominant for any type to play a strategy such that $S_i(h) = a_2^*$ and y is a possible outcome; by similar logic, $x' \rhd_i x$ implies $x' \notin P_i(h_2^*)$. Since we also know that $y \notin P_i(h_1^*)$, we can further write $y' \trianglerighteq_i y$ implies $y' \notin P_i(h_1^*)$; similarly, $x' \trianglerighteq_i x$ implies $x' \notin P_i(h_2^*)$.

Next, we show that for all types of agent i that reach h, it must be that $Top(\succ_i, P_i(h)) \neq x, y$. To start, assume that there is a type that reaches h such that $Top(\succ_i, P_i(h)) = x$. Let $y' \in P_i(h)$ be such that $y' \trianglerighteq_i y$ and y' is untrumped in $P_i(h)$, where it may be that $y' = y$ if y

[20] Note that y and z must both be undominated in $P_i(h)$—there is no $w \in P_i(h)$ such that $w \rhd_i y$ or $w \rhd_i z$—or else it would be impossible for them to be the top-ranked payoffs for types $\succ_i$ and $\succ_i'$, respectively. This, combined with richness, means that we can find such a type.

itself is untrumped. Consider a type such that $Top(\succ_i, P_i(h) \setminus \{x\}) = y'$, and that continues by ranking every y'' such that $y'' \trianglerighteq_i y$ immediately following y', in a way that is consistent with $\trianglerighteq_i$. Note that this type has no obviously dominant action at h: first, no action can obviously dominate a_1^*, since $x \notin G_i(h)$. Second, a_1^* itself cannot be obviously dominant. This is because, the worst case from a_1^* is strictly worse than y (since $x \notin G_i(h)$ and $y' \trianglerighteq_i y$ implies $y' \notin P_i(h_1^*)$), while $y \in P_i(h_2^*)$. Thus, this type has no obviously dominant action at h, which is a contradiction. Therefore, $Top(\succ_i, P_i(h)) \neq x$ for any type $\succ_i$ that reaches h; an analogous argument shows that $Top(\succ_i, P_i(h)) \neq y$ for any type that reaches h as well.

Thus, for all types that reach h, it must be that $Top(\succ_i, P_i(h)) \neq x, y$; further, by pruning, some such type is playing a strategy such that $S_i(\succ_i)(h) = a_1^*$ and x is a possible payoff. Let $Top(\succ_i, P_i(h)) = z$ for this type. The fact that $S_i(\succ_i)(h) = a_1^*$ implies that $z \in P_i(h_1^*)$ and $z \notin G_i(h)$ (if either were false, then it would not obviously dominant for this type to play a strategy such that $S_i(\succ_i)(h) = a_1^*$ and x is a possible payoff); in other words, $z \in PnG_i(h)$, and $z \in P_i(h_1^*)$. Since we just showed that $PnG_i(h) \cap P_i(h_1^*) \cap P_i(h_2^*) = \varnothing$, we have $z \notin P_i(h_2^*)$. Further, z is untrumped in $P_i(h)$ by construction, so $y \not\trianglerighteq_i z$; further, $z \not\trianglerighteq_i y$, or else it would not be obviously dominant for any type to follow a strategy such that $S_i(h) = a_2^*$ and y is a possible outcome. Let $y' \in P_i(h)$ be such that $y' \trianglerighteq_i y$ and y' is untrumped in $P_i(h)$, where it may be that $y' = y$ if y itself is untrumped. Consider a type $\succ_i$ such that $Top(\succ_i, P_i(h)) = z$ and $Top(\succ_i, P_i(h) \setminus \{z\}) = y'$, and that continues by ranking every y'' such that $y' \trianglerighteq_i y'' \trianglerighteq_i y$ immediately after y'. This type has no obviously dominant action at h: since $z \notin G_i(h)$ and $z \in P_i(h_1^*)$, no action $a \neq a_1^*$ can obviously dominate a_1^*; however, the worst case from a_1^* is strictly worse than y (since $z \notin G_i(h)$ and $y' \trianglerighteq_i y$ implies $y' \notin P_i(h_1^*)$), while $y \in P_i(h_2^*)$, and so a_1^* itself is also not obviously dominant, which is a contradiction. ■

Now, taken any OSP mechanism $(\Gamma, S_{\mathcal{N}})$ with perfect information such that Nature moves once, as the first mover. Further, prune this game according to the obviously dominant strategy profile $S_{\mathcal{N}} = (S_i(\succ_i))_{i \in \mathcal{N}}$. With slight abuse of notation, we denote this pruned, perfect information mechanism by $(\Gamma, S_{\mathcal{N}})$. By Claim 1, for each h, 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^* at h guarantees the best possible payoff in $P_i(h)$ for type $\succ_i$.[21] Let

$$\mathcal{S}_i(h) = \{S_i' : S_i'(h) \neq a^* \text{ and } S_i' \text{ is the obviously dominant strategy of some type } \succ_i \text{ that reaches } h\}.$$

[21]Consider a type $\succ_i$ such that the best possible payoff in $P_i(h)$ is x. Since type $\succ_i$ is choosing $S_i(h) = a' \neq a^*$, we must have $x \in P_i((h,a'))$, and, by Claim 1, x is guaranteeable at h. If type $\succ_i$ was playing a strategy S_i that did not guarantee x, this strategy would not be obviously dominant (because there is some other strategy, S_i', that does guarantee x).

By definition, each $S_i' \in \mathcal{S}_i(h)$ guarantees a unique payoff for i if she plays strategy S_i' starting from history h, no matter what the other agents do. 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 $P_i((h, a^*)) \notin G_i(h)$ 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 from the history (h', a'), where $a' = S_i^m(h')$ is the action that i would have played at h' in the original game had she followed strategy $S_i^m(\cdot)$. In other words, if i's strategy was 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.

The remainder of the proof of the lemma follows the argument in Pycia and Troyan (2023) and is included for completeness. 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 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 outcome 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, and i never moves again following a clinching action. This shows parts (i) and (ii). Part (iii) follows immediately from part (ii), due to greedy strategies and the pruning principle. ■

The above lemma shows that every OSP game is equivalent to one with at most one passing action at each history, which is Part 2 of Definition 1. The next lemma shows Part 3 of Definition 1, which deals with the payoff guarantees when an agent passes and some possible payoffs become impossible the next time she moves.

Lemma 2. *Let $(\Gamma, S_{\mathcal{N}})$ be a obviously strategy-proof mechanism that satisfies the conclusions of Lemma 1. At all h, if there exists a previously unclinchable payoff z that becomes impossible for agent i_h at h, then for all $x \in C^{\complement}_{i_h}(h)$: (i) if h is terminal, then $x' \trianglerighteq_i x$, where x' is the payoff for i that obtains at h, or (ii) if h is a history where agent i_h moves, then there exists an action $a \in A(h)$ such that for all $x' \in P_{i_h}((h,a))$, we have $x' \trianglerighteq_i x$.*

Proof. We start by defining an additional piece of notational that will be useful. Let

$$\bar{G}(S_i, h') = \{x : x \text{ is a payoff for } i \text{ at some terminal } \bar{h} \supseteq h' \text{ consistent with } i \text{ following } S_i\}.$$

This generalizes the notion of a guaranteeable payoff to a guaranteeable set of payoffs. In other words, by following strategy S_i starting at h', i can guarantee that her payoff will be something in the set $\bar{G}(S_i, h')$, but cannot guarantee which particular payoff in this set she will receive; note that if $x \in G_i(h')$, then there is some strategy S_i such that $\bar{G}(S_i, h') = \{x\}$ (indeed, by Lemma 1, x is not only guaranteeable, but also clinchable at h', and so any strategy such that $S_i(h')$ clinches x will have $\bar{G}(S_i, h') = \{x\}$).

Now, let h be any earliest history where some agent i moves such that there is a previously unclinchable payoff z that becomes impossible at h (the case for terminal histories will be dealt with separately below). This means that i moves at some strict subhistory $h' \subsetneq h$, and the following are true:

(a') For all $z' \trianglerighteq_i z$, $z' \notin P_i(h)$;

(b') For all $h' \subsetneq h$ such that $i_{h'} = i$, there exists some $z' \in P_i(h')$ such that $z' \trianglerighteq_i z$;

(c') For all $z' \trianglerighteq_i z$, $z' \notin C^{\complement}_i(h)$.

Take some $x \in C_i^c(h)$. If there exists a clinching action $a \in A(h)$ that clinches some $x' \trianglerighteq_i x$ for i, then we are done. Thus, assume towards a contradiction that there exists some payoff $x \in C_i^c(h)$ such that there is no clinching action $a \in A(h)$ that clinches some $x' \trianglerighteq_i x$ for i. Note that this, combined with Lemma 1, part (ii) implies the following:

(*) For any $x' \trianglerighteq_i x$, we have $x' \notin G_i(h)$.

Consider a type $\succ_i$: $z_1, z_2 \ldots, z, \ldots$, where $\{z_1, z_2, \cdots, z\}$ is the set of all of the payoffs that $\trianglerighteq_i$-dominate z, and the notation means that type $\succ_i$ ranks all payoffs in this set higher than any payoff not in the set, in some way that is consistent with $\trianglerighteq_i$. By (b') and (c'), any obviously dominant strategy must have type $\succ_i$ passing at all $h' \subsetneq h$ where she is called to move.

Claim 2. There must exist a strategy $\tilde{S}_i$ such that for any $x' \in \bar{G}(\tilde{S}_i, h)$, we have $x' \trianglerighteq_i x$. Further, for this strategy, we have $|\bar{G}(\tilde{S}_i, h)| \geq 2$.

Proof of claim. To show the first part, assume to the contrary that for all S_i', there exists some $y \in \bar{G}(S_i', h)$ such that $y \ntrianglerighteq_i x$. Consider a type $\succ_i$: $z_1, z_2 \ldots, z, x_1, x_2, \ldots, x, \ldots$, where $\{z_1, z_2, \ldots, z\}$ is the set of all payoffs that $\trianglerighteq_i$ −dominate z, and $\{x_1, x_2, \ldots, x\}$ is the set of all payoffs that $\trianglerighteq_i$ −dominate x and not z.[22] At any $h' \subsetneq h$, any obviously dominant strategy must have this type passing. Now, consider h itself. For any $y \in \bar{G}(S_i', h)$, the worst case outcome for type $\succ_i$ from S_i' starting from h is some payoff that is weakly worse than y. Further, there exists a $y \in \bar{G}(S_i', h)$ that is strictly worse than x, which follows from $y \ntrianglerighteq_i x$ and $y \ntrianglerighteq_i z$, where the latter is because $z_1, z_2, \ldots, z \notin P_i(h)$, by (a'). However, we also have $x \in C_i(h')$ for some $h' \subsetneq h$, and so, the best case outcome from clinching x at h' is x. This implies that type $\succ_i$ does not have an obviously dominant strategy, a contradiction.

To show that $|\bar{G}(\tilde{S}_i, h)| \geq 2$, assume that $|\bar{G}(\tilde{S}_i, h)| = 1$, which means that $\bar{G}(\tilde{S}_i, h) = \{x'\}$ for some $x' \trianglerighteq_i x$. This implies that $x' \in G_i(h)$, which contradicts (*), and so we conclude that, $|\bar{G}(\tilde{S}_i, h)| \geq 2$. ■

The previous claim shows that if x was previously clinchable and nothing dominating it is clinchable at h, then there at least must be a strategy that, starting from h, guarantees some set of payoffs that all weakly dominate x. The next claim shows that at h, there must exist a passing action, and further, all payoffs that are possible after passing must weakly dominate x.

Claim 3. There must be a passing action $a^* \in A(h)$, and for all $x' \in P_i((h, a^*))$, we have $x' \trianglerighteq_i x$.

[22]Note that $x \ntrianglerighteq_i z$, by condition (c'), and so $x \notin \{z_1, z_2, \ldots, z\}$. However, there may be payoffs that dominate both z and x; such payoffs are included in $\{z_1, z_2, \ldots, z\}$.

Proof of claim. Let $\tilde{S}_i$ be the strategy in the statement of Claim 2, and note that for any $x' \in \bar{G}(\tilde{S}_i, h)$, we have $x' \notin C_i(h)$, by (*). Therefore, x' can only be possible following a passing action at h, and so such an action a^* must exist.

It remains to prove the second part of the claim. By way of contradiction, assume that there is some $y \in P_i((h, a^*))$ such that $y \ntrianglerighteq_i x$. First, note $x \ntrianglerighteq_i y$. To see why, note that if $x \trianglerighteq_i y$, then $y \ntrianglerighteq_i x$ and $x \trianglerighteq_i y$ implies $x \succ_i y$ for all types $\succ_i$ of agent i. Because the game is pruned, there exists a type $\succ_i$ such that outcome y is on-path for this type's OSP strategy $S_i(\succ_i)$. However, i could have clinched x at some history $h' \subset h$, which contradicts that $S_i(\succ_i)$ is obviously dominant. Thus, $x \ntrianglerighteq_i y$. Further, since $y \ntrianglerighteq_i x$, we have $y \notin \bar{G}(\tilde{S}_i, h)$, by Claim 2.

Now, let $\bar{x}$ be a payoff that is undominated in $\bar{G}(\tilde{S}_i, h)$, i.e., for any other $x' \in \bar{G}(\tilde{S}_i, h)$, $x' \ntrianglerighteq_i \bar{x}$. Note also that for all $x' \in \bar{G}(\tilde{S}_i, h)$, $y \ntrianglerighteq_i x'$ since otherwise, $y \trianglerighteq_i x'$ and $x' \trianglerighteq_i x$ would imply $y \trianglerighteq_i x$. By richness, there exists a type such that $\succ_i$: $z_1, z_2, \cdots, z, \cdots, \bar{x}, y, \cdots, x, \cdots$; in words, type $\succ_i$ ranks z and everything that dominates it highest, followed by everything that dominates $\bar{x}$ but not z (which are therefore not possible at h), followed by immediately by y, followed by everything else.

Let $S_i(\succ_i)$ be this type's obviously dominant strategy. Note that for any $h' \subseteq h$, $S_i(\succ_i)$ must be the passing action. This includes at h itself, by (*). Also by (*) and the construction of $\succ_i$, the worst case from $S_i(\succ_i)$ for type $\succ_i$ starting from h is strictly worse than $\bar{x}$. If $S_i(\succ_i) \neq \tilde{S}_i$, then $S_i(\succ_i)$ is not obviously dominant, since $\bar{x}$ is possible from $\tilde{S}_i$. Thus, the only remaining possibility is that is $S_i(\succ_i) = \tilde{S}_i$. In this case, the worst case from $S_i(\succ_i)$ is strictly worse than y. This follows because in Claim 2, we showed that $\bar{G}(\tilde{S}_i, h)$ has at least two elements, and by construction of $\succ_i$, one of these must be strictly worse than $\bar{x}$, and, since $y \notin \bar{G}(\tilde{S}_i, h)$, also strictly worse than y. On the other hand, since $y \in P_i((h, a^*))$, there is some other strategy $S_i' \neq \tilde{S}_i$ where y is an on-path outcome if i follows S_i' starting from h. But this means that $S_i(\succ_i) = \tilde{S}_i$ is not obviously dominant for type $\succ_i$, either, a contradiction. ■

The previous claim completes the argument for all nonterminal histories at which an agent i moves. Last, we must consider the case where h is a terminal history, $h = \bar{h}$. As above, let z be a payoff such that (a'), (b'), and (c') hold (replacing h with $\bar{h}$). Recall that for terminal histories, we define $C_i(\bar{h}) = \{y\}$ for all i, where y is the payoff that obtains at $\bar{h}$. Towards a contradiction, assume that there exists some $h' \subsetneq \bar{h}$ such that $i_{h'} = i$ and some $x \in C_i(h')$ such that $y \ntrianglerighteq_i x$. Note that we also have $y \ntrianglerighteq_i z$ (by (a')) and $x \ntrianglerighteq_i z$ (by (c')). In other words, x, y, z must all be distinct payoffs for i such that $y \ntriangleright_i z$, $x \ntriangleright_i z$, and $y \ntriangleright_i x$. Thus, by richness, there exists a type such that $z \succ_i x \succ_i y$ (there may be other payoffs ranked above z and in between z, x, and y). By (b') and (c'), there is some $z' \trianglerighteq_i z$ possible at every

$h \subsetneq \bar{h}$ where i is to move, but no $z' \trianglerighteq_i z$ is 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 this strategy is not obviously dominant (because outcome y is on-path from the proposed strategy), which is a contradiction. ■

Lemmas 1 and 2 complete the proof of Proposition 1. ■

Proposition 2. *Every generalized millipede mechanism is OSP.*

Proof of Proposition 2. Let Γ be a millipede game. Let $Top(\succ_i, X')$ denote the best possible payoff in the set X' according to preferences $\succ_i$. Using this notation, a strategy $S_i(\succ_i)$ is a *greedy strategy* if, whenever $Top(\succ_i, C_i(h)) = Top(\succ_i, P_i(h))$, the action $S_i(\succ_i)(h)$ clinches the top payoff in $P_i(h)$, and if $Top(\succ_i, C_i(h)) \neq Top(\succ_i, P_i(h))$, then $S_i(\succ_i)(h)$ passes at h.[23]

Consider some profile of greedy strategies $(S_i(\cdot))_{i \in \mathcal{N}}$. It is clear that if $Top(\succ_i, C_i(h)) = Top(\succ_i, P_i(h))$, then 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.

Assume that there exists a history h that is on the path of play for type $\succ_i$ when following $S_i(\succ_i)$ such that $Top(\succ_i, C_i(h)) \neq Top(\succ_i, P_i(h))$, 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 payoff from passing (and continuing to follow $S_i(\succ_i)$ at all future nodes).

First, note that $x_W(h) \succsim_i x_W(h')$ for all $h' \subsetneq h$ such that $i_{h'} = i$. Since passing is obviously dominant at all $h' \subsetneq h$, 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 such h'. At h, since passing is not obviously dominant and all other actions are clinching actions, we have $x_C(h) \succ_i x_W(h)$; further, since $Top(\succ_i, C_i(h)) \neq Top(\succ_i, P_i(h))$, there must be some $x' \in P_i(h) \setminus C_i(h)$ such that $x' \succ_i x_C(h) \succ_i x_W(h)$. The above implies that $x' \succ_i x_C(h) \succ_i x_C(h')$ for all $h' \subsetneq h$ such that $i_{h'} = i$.

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 $x_1 \succ_i x_2 \succ_i \cdots \succ_i x_M$.

Consider a path of play starting from h that is consistent with $S_i(\succ_i)$ and ends in a terminal history $\bar{h}$ at which i receives $x_W(h)$. For every $x_m \in X_0$, let h_m denote the earliest history on this path such that $x_m \notin P_i(h_m)$ and either (i) $i_h = i$ or (ii) h_m is terminal. Note that because i is ultimately receiving payoff $x_W(h)$, such a history h_m exists for all $x_m \in X_0$.

[23]There may be multiple ways for i to clinch payoff x at h. Further, x may still be possible if i passes at h.

Let $\hat{h} = \max\{h_1, h_2, \ldots, h_M\}$ (ordered by $\subset$); in words, $\hat{h}$ is the earliest history at which no payoffs in X_0 are possible any longer. 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.

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

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 $P_i(h') \subseteq P_i(h)$, 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 $x_W(h)$.[24]

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 $x_{m'} \succ_i x_m$ be (a) payoff that becomes impossible at $\hat{h}_{-m}$. There are two cases:

Case (i): $h' \subset \hat{h}_{-m}$. This is the case in which x_m is clinchable while there is some strictly preferred payoff $x_{m'} \succ_i x_m$ that is still possible. Since $x_{m'}$ is (by definition) the last payoff in $\{x_1, \ldots, x_{m-1}\}$ to become impossible, we can conclude that $x_{m'}$ becomes impossible at $\hat{h}_{-m}$. Further, by the induction step, all $\{x_1, \ldots, x_{m-1}\}$ are previously unclinchable at $\hat{h}_{-m}$, and so $x_{m'}$ is previously unclinchable at $\hat{h}_{-m}$. Thus, by point 3 in the definition of a millipede game, there must be some action $a \in A(\hat{h}_{-m})$ such that for all $y \in P_i((\hat{h}_{-m}, a))$, $y \trianglerighteq_i x_m$. Since all payoffs strictly preferred to x_m are no longer possible at $\hat{h}_{-m}$, we must have $y = x_m$, and x_m is clinchable at $\hat{h}_{-m}$. Thus, x_m is the best remaining payoff at $\hat{h}_{-m}$, and is clinchable, and so greedy strategies direct i to clinch x_m at $\hat{h}_{-m}$, which contradicts that she receives $x_W(h)$.[25]

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, let $\hat{h} = \max\{h_1, h_2, \ldots, h_M\}$ and let $\hat{x}$ be a payoff that becomes impossible at $\hat{h}$. Since $\hat{x}$ is the last payoff in X_0 to become impossible, we can conclude that $\hat{x}$ becomes impossible at $\hat{h}$. Further, the claim shows that no $x \in X_0$ is clinchable at any $h' \subseteq \hat{h}$, and so we can further conclude that $\hat{x}$ is previously unclinchable at $\hat{h}$. Therefore, by part 3 in the definition of a millipede game, there is some action $a \in A(\hat{h})$ such that for all $y \in P_i((\hat{h}, a))$, $y \trianglerighteq_i x_C(h)$. Since all preferred payoffs are no longer possible at $\hat{h}$, we must have $y = x_C(h)$ and greedy strategies direct i to clinch $x_C(h)$, which contradicts that she

[24] Recall that for terminal histories h, we define $C_i(h) = \{x\}$, where x is the 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 $x_W(h)$.

[25] Note that if $\hat{h}_{-m}$ is a terminal history, the argument still applies, and the outcome associated with this terminal history must be x_m, which contradicts that i gets payoff $x_W(h)$.

receives $x_W(h)$.[26] ■

Propositions 1 and 2 complete the proof of Theorem 1. ■

[26] If $\hat{h}$ is a terminal history, then we make an argument analogous to footnote 24 to reach the same contradiction.