Are Simple Mechanisms Optimal when Agents are Unsophisticated?*

Jiangtao Li ⓡ Piotr Dworczak†

March 13, 2024

Abstract

We study the design of mechanisms involving agents that have limited strategic sophistication. We define a mechanism to be *simple* if—given the assumed level of strategic sophistication—agents can determine their optimal strategy. We examine whether it is *optimal* for the mechanism designer who faces strategically unsophisticated agents to offer a simple mechanism. We show that when the designer uses a mechanism that is not simple, while she loses the ability to predict play, she may nevertheless be better off no matter how agents resolve their strategic confusion.

KEYWORDS: simplicity, complexity, strategic sophistication, robust mechanism design, obviously strategy-proof mechanisms

---

*We thank Mohammad Akbarpour, Tilman Börgers, Gabriel Carroll, Jeff Ely, Scott Duke Kominers, Shengwu Li, Paul Milgrom, Stephen Morris, Parag Pathak, Andrzej Skrzypacz, Rani Spiegler, Satoru Takahashi, Takuro Yamashita, and participants at various seminars and conferences for helpful discussions. An abstract of this paper appeared in the *Proceedings of the 22nd ACM Conference on Economics and Computation (EC'21).* Li thanks the Department of Economics at the Northwestern University for hospitality and support during the academic year 2018-2019, when some of the work on this paper was undertaken.

†The "ⓡ" symbol indicates that the authors' names are in certified random order, as described by Ray ⓡ Robson (2018). *Li:* School of Economics, Singapore Management University; jtli@smu.edu.sg. *Dworczak:* Department of Economics, Northwestern University; Group for Research in Applied Economics; piotr.dworczak@northwestern.edu.

1 Introduction

It is widely accepted that "real-life" economic agents are not as rational as their counterparts in economic models. When agents have limited strategic sophistication, economists lose confidence in the performance of mechanisms that require participants to engage in complicated mental tasks. For example, achieving a Bayesian Nash equilibrium requires each agent to know the distribution of their opponents' private information and correctly forecast the strategies they play; this is why dominant-strategy (or strategy-proof or SP) mechanisms are generally perceived to be superior for practical purposes. Following the mounting evidence that even dominant strategies are difficult to identify for real-life agents, several recent papers introduced classes of mechanisms in which agents can determine their optimal strategy under even weaker assumptions about strategic sophistication. For instance, Li (2017) proposes the notion of obviously strategy-proof (OSP) mechanisms in which agents can determine their optimal strategy even if they cannot engage in contingent reasoning.

This paper studies the design of mechanisms involving agents that have limited strategic sophistication. We call a mechanism *simple* if, given the assumed level of strategic sophistication, agents can determine their optimal strategy in the mechanism. For example, if we are only comfortable assuming that agents avoid obviously dominated strategies (Li, 2017), then an OSP mechanism is simple because the obviously dominant strategy is the unique strategy that is not obviously dominated. If the designer instead offers a mechanism that is not OSP, she can no longer predict how agents will behave. We call a mechanism *complex* if it creates *strategic confusion* for the agents, understood as the inability to determine their optimal strategy.

We introduce a general framework for studying simple and complex mechanisms. We use a black-box approach to modeling strategic sophistication by working with arbitrary partial orders over strategies available to each player. A strategy is ranked higher than another strategy if the agent choosing between them can recognize that the former strategy is superior. A mechanism is simple if each agent has a unique strategy that is maximal according to the partial order; a mechanism is complex otherwise. This framework, while abstract, allows us to identify economic trade-offs that underlie the choice between simple and complex mechanisms for a wide range of solution concepts.

Our key observation is that the inability of the designer to predict the outcome of a complex mechanism need not be a sufficient reason to use simple mechanisms. As long as the designer is ultimately concerned with maximizing her own payoff—which is typically assumed in mechanism design—complex mechanisms may in fact be strictly preferred by the designer to simple ones. The following example illustrates our point.

Example 1. In this and the next example, we consider agents that avoid obviously dominated strategies but do not make any assumption about how agents choose among the remaining strategies.[1] OSP mechanisms are simple, while non-OSP mechanisms are complex—their outcomes cannot be uniquely predicted.

Consider the problem of designing a trading platform for two traders, A and B, with the goal of maximizing the platform's intermediation profit. Each trader can buy or (short) sell one unit of the asset. Trader A's valuation for the asset is either 0 or 2/3. Trader B's valuation for the asset is either 1/3 or 1. The designer believes individual types to be equally likely but correlated across traders: $\pi((0,\,1/3)) = \pi((2/3,\,1)) = 2/5$.[2] The platform cannot hold inventory, so we impose ex-post market-clearing.

The optimal OSP mechanism yields an expected profit of 1/5 for the platform and is implemented by the following extensive-form game, which can be viewed as an ascending personal-clock auction:

1. Trader A is asked whether she would like to sell the asset at the price 0; if she says "yes," then that trade is implemented; if she says "no," then:
2. Trader B is asked whether she would like to sell the asset at the price 1/3; if she says "yes," then that trade is implemented; if she says "no," then there is no trade.

Conditional on trade, the platform charges a fee of 1/3 to the buyer, that is, the buyer pays the trading price plus 1/3.[3]

It is obviously dominant for type 0 of trader A to accept the offer in the first stage, and for type 2/3 to reject it. It is obviously dominant for type 1/3 of trader B to accept

[1] A strategy obviously dominates another strategy if, at any information set where the two strategies first diverge, the best outcome under the second strategy is no better than the worst outcome under the first strategy. A strategy is obviously dominated if there exists another strategy that obviously dominates it. A strategy is obviously dominant if it obviously dominates any other strategy. A mechanism is OSP if it has an equilibrium in obviously dominant strategies. See Appendix A.1 for formal definitions.

[2] Because each trader's role as a buyer or a seller is endogenously determined, the binding incentive constraints cannot be pinned down ex-ante. The correlation of types (and the specific value of 2/5) plays no role in our analysis, except for ensuring that the type profile (0, 1) is relatively unlikely, which leads to a particular structure of binding IC and IR constraints in the optimal simple mechanism. For related models, see Cramton et al. (1987), Lu and Robert (2001), Chen and Li (2018), and Loertscher and Marx (2020).

[3] We show the optimality of this mechanism in Appendix A.2.

the offer in the second stage, and for type 1 to reject it. It follows that the platform's profit is 1/3 except when the type profile is (2/3, 1). Intuitively, the inefficient no-trade outcome at the type profile (2/3, 1) is implemented so that type 0 has an obviously dominant strategy: If trader B were to buy the asset from trader A conditional on the profile (2/3, 1), then the best possible outcome for type 0 from rejecting the initial offer would yield a strictly positive payoff, while her equilibrium strategy yields a payoff of 0.

Consider now an alternative mechanism:

1. Trader A is asked whether she would like to sell the asset at the price 0; if she says "yes," then that trade is implemented; if she says "no," then:
2. Trader B is asked whether she would like to sell the asset at the price 1/3; if she says "yes," then that trade is implemented; if she says "no," then:
3. Trader A is asked whether she would like to sell the asset at the price 2/3; if she says "yes," then that trade is implemented; if she says "no," then there is no trade.

Conditional on trade, the platform charges a fee of 1/3 to the buyer.

In comparison to the optimal OSP mechanism, this alternative mechanism gives trader A an option of selling the asset at the price 2/3 should the two traders fail to reach an agreement in the first two stages. In particular, trade happens at the type profile $(2/3, 1)$ because it is obviously dominant for type 2/3 of trader A to reject the first offer but to accept the second one. However, as discussed above, the best possible outcome for type 0 from rejecting the first offer now yields a strictly positive payoff. Thus, this mechanism is no longer OSP: Type 0 of trader A is confused between accepting the initial offer (which gives her 0) and rejecting it while accepting the second offer (which gives her $-2/3$ or 2/3, depending on the behavior of trader B).[4] The key observation is that, regardless of how trader A resolves this confusion, trade always happens. Thus, the platform achieves a profit of 1/3 ex post, and hence also in expectation. Finally, for each type, non-participation is obviously dominated.[5] Thus, as long as traders do not play obviously dominated strategies, by adopting a complex mechanism, the platform is guaranteed to achieve a *strictly* higher revenue than in the optimal OSP mechanism. □

In Example 1, there exists a complex mechanism that generates a strictly higher expected payoff for the designer than the best simple mechanism *regardless* of how agents

[4]Both types of trader B still have obviously dominant strategies: It is obviously dominant for type 1/3 to accept her offer, and for type 1 to reject it.

[5]The platform could ensure that non-participation is obviously *strictly* dominated by adjusting the fee and the prices by an arbitrarily small $\epsilon > 0$.

resolve their strategic confusion. In such cases, we say that the complex mechanism *strongly dominates* the best simple mechanism.

We derive two general results concerning strong dominance. These results provide sufficient conditions under which, respectively, the best simple mechanism may or may not be strongly dominated. For the first result, we observe that simple mechanisms may be overly restrictive by requiring that no type be strategically confused. In many settings, the set of outcomes implementable by simple mechanisms shrinks as the type space expands. In such cases, the designer faces a trade-off: Simplicity of the mechanism (for all types) comes at the cost of excluding the possibility of implementing certain outcomes conditional on a subset of types. Thus, the designer might benefit from imposing simplicity only for a subset of types with the remaining types being potentially confused. We formalize this idea by defining a property called "accommodation of additional types" (AAT). The AAT property requires that, for any simple mechanism on a subset of the type space, we can accommodate additional types while implementing the same outcome on the subset. We show that whenever the AAT property is violated, strong dominance of the best simple mechanism occurs for some objective function of the designer. The AAT property fails in many classical social choice environments. We illustrate this via a voting example in which the designer attempts to maximize Rawlsian welfare of two agents with privately observed preferences over three alternatives. For the solution concept of OSP, the best simple mechanism—dictatorship—is strongly dominated by a complex mechanism in which both agents influence the final choice of the alternative.

For the second result, we identify environments in which simple mechanisms are *not* strongly dominated. We show that the optimal simple mechanism is not strongly dominated when the designer's maximized objective function is the same as the value of a relaxed problem in which the incentives constraints are only imposed along the edges of some directed tree in the type space. A notable instance of such a setting under the solution concept of OSP is the single-unit auction. In this and other settings, the result establishes an *optimality foundation* for the use of simple mechanisms.

The notion of strong dominance is demanding, as it requires that the superior complex mechanism generate a *strictly higher* expected payoff to the designer, even if agents choose their strategies to minimize the designer's expected payoff whenever they are confused. This property ensures robustness to how agents choose among undominated strategies; if agents turn out to be more sophisticated and rule out more strategies (than hypothesized by the designer), the performance guarantee of a complex mechanism can only improve. However, in some applications, the designer may be satisfied with a weaker property of a superior complex mechanism: It *never* yields a lower payoff than the best simple mechanism, and *sometimes* yields a strictly higher payoff. We call this property "weak dominance."[6]

By means of an extensive example focusing on the solution concept of OSP, we argue that weak dominance of the best simple mechanism is a common phenomenon. Intuitively, a weakly dominant complex mechanism can often be constructed by augmenting the simple mechanism with an additional option that—if taken—benefits the designer. The additional option is made sufficiently attractive for the agent so that choosing it cannot be ruled out given the assumed level of sophistication. The following example illustrates this logic.

Example 2. Consider the problem of selling an item to one of N bidders with independent private values distributed according to a regular and symmetric distribution. The best simple mechanism is an ascending clock auction with a distribution-dependent reserve price. In the ascending clock auction, active bidders choose whether to exit as the clock price increases, and bidders who exit remain inactive thereafter. The auction stops when all but one bidder exit. The remaining bidder wins the object and pays the clock price.

The following modified mechanism, which we call the ascending clock auction with jump bidding, weakly dominates the best simple mechanism: Each bidder is allowed to speed up the clock by jump bidding, that is, to make a higher bid than the current clock price. This mechanism is not OSP because making a jump bid (to a bid b) is not obviously dominated for a bidder with value $v > b$ at the clock price $p < b$. Indeed, making a jump bid to b yields the best-case payoff of $v - b$ to the bidder, while following the default strategy (of exiting when the clock price reaches v) yields a payoff of 0 in the worst case.[7] Agents are strategically confused as they now have multiple strategies that are not obviously dominated—the mechanism is complex.

Given the assumed strategic sophistication of the bidders, the designer cannot

[6]It may be useful to draw an analogy to the notion of weak dominance between two strategies in game theory. The arguments for and against playing weakly dominated strategies carry over to selecting weakly dominated mechanisms.

[7]Of course, bidders would not jump bid if they could engage in contingent reasoning—jump bidding is weakly dominated but not obviously dominated.

predict whether jump bidding will occur or not. Nonetheless, a revenue-maximizing auctioneer might prefer the ascending clock auction with jump bidding to the ascending clock auction: If none of the bidders jump bids, then the performance of the ascending clock auction with jump bidding is the same as that of the ascending clock auction; in the event that some bidder jump bids, the expected revenue of the ascending clock auction with jump bidding is strictly higher than that of the ascending clock auction. □

We use a similar construction to show that the best OSP mechanism is weakly dominated for a revenue-maximizing designer in a variety of settings. On a flip side, we also prove that, generically, single-agent posted price mechanisms are *not* weakly dominated.

Overall, the paper proposes a systematic framework for thinking about the issues of strategic simplicity and complexity. By offering results that both oppose and support the use of simple mechanisms, we emphasize that whether simplicity is desirable or not is not merely a function of agents' sophistication. Instead, each environment in question must be carefully analyzed before deciding whether a simple mechanism should be used.

1.1 Related literature

Simple mechanisms: Our paper contributes to the litrature on the design of mechanisms involving strategically unsophisticated agents. While Li (2017), Börgers and Li (2019), and Pycia and Troyan (2023) provide notions of simplicity and characterize simple mechanisms according to these notions, our focus is on the trade-off between simplicity and optimality, and on whether there is a foundation for the use of simple mechanisms from an optimality perspective.[8] Indeed, we show that in many cases, the designer might prefer a mechanism that is not simple.

Robust mechanism design: Traditional models in mechanism design make strong assumptions about the detailed knowledge of the designer about the inputs to the mechanism design model. The literature of robust mechanism design seeks to relax these assumptions; see Carroll (2019) for a recent survey. While the leading interpretation of our exercise is that agents have limited strategic sophistication, an alternative interpretation—tying our work to the burgeoning literature of robust mechanism design—is that we

[8]There are, of course, many dimensions of simplicity. Our paper and the papers cited here consider the strategic dimension.

relax the assumption about the designer's knowledge of the strategic reasoning process of the agents (beyond some minimal rationality assumptions).

Implementation in undominated strategies: To the best of our knowledge, this paper is the first to study the trade-off between simplicity and optimality under the black-box approach to solution concepts modeled via the partial order on strategies; it is also first to provide examples of strong and weak dominance for the solution concept of OSP, which is our leading application. However, for the solution concept of SP, our analysis has antecedents in the literature that study implementation in undominated strategies; see, for example, Börgers (1991), Jackson (1992), Börgers and Smith (2012), Carroll (2014), Yamashita (2015), Mukherjee et al. (2019), and Mukherjee et al. (2024). In particular, it is known from these papers that the optimal SP mechanism could be weakly or strongly dominated by complex mechanisms; our analysis contributes new results and examples, and extends some of the insights from this literature to other solution concepts.

Simple versus complex mechanisms: There is a large computer-science literature—too vast to be surveyed here—that quantifies the worst-case loss from using certain simple classes of mechanisms relative to the optimal mechanism; for example, see Hartline and Roughgarden (2009). Pycia and Troyan (2023) ask whether restricting attention to simple mechanisms entails any loss by studying how the set of allocations that are simply-implementable depends on agents' degree of strategic sophistication. The fundamental difference to our work is that these papers compare the performance of mechanisms *across* solution concepts and the corresponding classes of mechanisms. In contrast, we fix the assumption about agents' rationality and compare the performance of simple mechanisms (in which agents know what to do) with the performance of complex mechanisms (in which agents are confused).

Our paper is related to the growing literature that studies bounded rationality in industrial organization and shows that firms might benefit from purposefully confusing consumers in the presence of naive consumers; see Spiegler (2011) and references therein. Glazer and Rubinstein (2012, 2014) study persuasion and mechanism design models with boundedly rational agents and show that a listener could benefit from committing to a complex mechanism that makes it difficult for a dishonest speaker to cheat. Jakobsen (2020) considers a boundedly rational agent who has both imperfect memory and limited deductive (computational) ability, and shows that the principal could benefit by using a complex contract. These papers impose specific assumptions about the reasoning procedure of the agent (even in complex mechanisms). In contrast, we take a robust approach, and do not impose any assumptions on how agents resolve their confusion in complex mechanisms.

2 Framework

We first introduce a general framework for studying simple and complex mechanisms. We model the strategic reasoning by a player via a partial order on the set of available strategies. We use the solution concept of obvious dominance as a leading example.

Environment. There is a finite set of N players, $\mathcal{N} = \{1, 2, \ldots, N\}$, and an arbitrary (possibly infinite) set of alternatives $\mathcal{X}$. Each player i has payoff-relevant information indexed by $\theta_i \in \Theta_i$, where Θ_i is finite. We refer to θ_i as player i's type. The set of possible type profiles is $\Theta = \times_{i \in \mathcal{N}} \Theta_i$ with its representative element $\theta = (\theta_1, \theta_2, \ldots, \theta_N)$. The type profile is distributed according to a prior probability distribution $\pi \in \Delta\Theta$.

Each player i is endowed with a utility function $u_i : \mathcal{X} \times \Theta_i \to \mathbb{R}$ (we assume private values). The designer has a utility function $v : \mathcal{X} \times \Theta \to \mathbb{R}$. That is, $u_i(x, \theta_i)$ and $v(x, \theta)$ denote type θ_i's utility and the designer's utility, respectively, when the type profile is θ and the implemented alternative is x. We assume that the designer knows the distribution π and is an expected utility maximizer.

Mechanisms. We consider mechanisms that are finite, imperfect-information, extensive-form games with perfect recall and consequences in $\mathcal{X}$, belonging to some arbitrarily large but fixed set $\mathcal{M}$.[9] We relegate the formal definition of an extensive-form game to Appendix A.1. For our current purposes, all that matters is that every game $\Gamma \in \mathcal{M}$ specifies a finite set of (pure) strategies $\mathcal{S}_i$ available to player i, with a representative element S_i.

In our abstract framework, a solution concept τ associates with every game $\Gamma \in \mathcal{M}$ and type θ_i of player i a partial order (a reflexive and transitive binary relation) $\preceq_{\theta_i}$ on the set of strategies $\mathcal{S}_i$ available to player i. We let $S_i \sim_{\theta_i} S_i'$ denote the equivalence

[9]Finiteness is an important assumption: Jackson (1992) shows that infinite mechanisms may implement virtually any decision rule in undominated strategies, relying on an infinite hierarchy of weakly dominated strategies with no dominant strategy "at the top."

relation associated with $\preceq_{\theta_i}$ (that is, $S_i \sim_{\theta_i} S'_i$ if both $S_i \preceq_{\theta_i} S'_i$ and $S'_i \preceq_{\theta_i} S_i$). We let $\prec_{\theta_i}$ denote the strict partial order induced by $\preceq_{\theta_i}$ (that is, $S_i \prec_{\theta_i} S'_i$ if $S_i \preceq_{\theta_i} S'_i$ but $S'_i \npreceq_{\theta_i} S_i$). The interpretation of $S_i \prec_{\theta_i} S'_i$ is that type θ_i of player i recognizes strategy S_i as unambiguously worse than strategy S'_i. Whenever $S_i \prec_{\theta_i} S'_i$, we say that S'_i *dominates* S_i (for type θ_i), or, equivalently, that S_i *is dominated by* S'_i (for type θ_i). The relation $S_i \sim_{\theta_i} S'_i$ is interpreted as the agent's *indifference* between the two strategies; in this case, we say that the two strategies are equivalent for type θ_i. Finally, if the two strategies S_i and S'_i are not ranked by $\preceq_{\theta_i}$, the agent is unsure how to make a choice between them.

Throughout, we rely on the straightforward interpretation of the partial order as capturing the agent's true (and typically incomplete) ranking over strategies that is known to the designer. However, a more precise interpretation of the model is that the order induced by τ reflects what the designer is willing to *assume* about the agent's induced preference over strategies. For example, even if the agent "knows" which of the strategies S_i or S'_i she prefers to choose, the partial order would not rank them if the designer lacked the knowledge needed to determine the agent's choice between S_i and S'_i.

In our leading example, the partial order $\prec_{\theta_i}$ captures (weak) obvious dominance, as introduced by Li (2017). Under this solution concept, a strategy S_i is dominated by a strategy S'_i for type θ_i of player i, if starting at any earliest point of departure, the worst-possible payoff for θ_i from playing S'_i (across all possible choices of other players) is weakly better than the best-possible payoff for θ_i (across all possible choices of other players). We provide formal definitions in Appendix A.1.

In general, a solution concept τ in our framework can be an arbitrary mapping from the set of games to a profile of partial orders on strategies. This flexibility allows us to cover a wide range of solution concepts that were studied in the literature (we discuss additional examples at the end of this section).[10]

Strategic confusion, simple and complex mechanisms. For any mechanism Γ, we let

$$U_i(\theta_i) = \{S_i \in \mathcal{S}_i : \nexists\, S'_i \in \mathcal{S}_i \text{ such that } S'_i \succ_{\theta_i} S_i\}.$$

[10]The flexibility also permits nonsensical solution concepts: no consistency is imposed on how the partial order varies across games (e.g., an agent can behave very differently in games that look arbitrarily similar) and how it relates to agents' payoffs (e.g, an agent may be minimizing their own payoff)—our results remain true in such cases but will have little or no economic significance.

That is, $U_i(\theta_i)$ is the set of strategies that are not dominated for type θ_i. We will call such strategies *undominated*. Given the finiteness of the mechanism, the set of undominated strategies is non-empty. The designer assumes that players will avoid dominated strategies:

Assumption 1 (Behavioral assumption)**.** *For any $i \in \mathcal{N}$, each type θ_i of player i plays a strategy from the set $U_i(\theta_i)$.*

The behavioral assumption allows us to interpret the partial order as reflecting the agent's sophistication. For example, a clueless agent might be unable to rank any of her strategies in any mechanism; based on Assumption 1, the designer cannot make any prediction about that agent's behavior. On the other extreme, a Bayesian, fully-rational agent with perfect insight into other agents' choices could be captured via a total order on her strategies induced by expected utility comparisons. The interesting cases are the intermediate ones in which the agent is able to rank some, but not necessarily all, strategies. For example, obvious dominance captures agents who are sophisticated enough to rule out obviously dominated strategies but not sophisticated enough to rule out strategies that are weakly dominated (in the usual sense) but not obviously dominated.[11]

Having fixed a solution concept τ, based on our behavioral assumption, we can classify mechanisms as being either simple or complex.

Definition 1 (Strategic confusion and complex mechanisms)**.** *Fixing a mechanism Γ, type θ_i of player i is strategically confused if $U_i(\theta_i)$ contains at least two strategies (that are not equivalent for type θ_i); in this case, we call mechanism Γ complex.*

Definition 2 (Simple mechanisms)**.** *A mechanism Γ is simple if for any player $i \in \mathcal{N}$, no type θ_i is strategically confused.*

Strategic confusion means that at least one type of some player has more than one undominated strategy in the mechanism, and therefore, in the absence of further assumptions on behavior or strategic reasoning, it is impossible to determine which strategy she will select. In contrast, a simple mechanism gives each type of every player a

[11] In this case, Assumption 1 is the same assumption that forms the basis for the simplicity notion of OSP. If one thinks that participants will play obviously dominant strategies when they are available, then one seems also compelled to reason that participants will avoid obviously dominated strategies (even when no strategy is obviously dominant).

unique (up to equivalence) undominated strategy. Such strategy must then be *dominant* in the sense that it dominates any other strategy.

Every solution concept τ is associated with a different class of simple mechanisms. For example, when the ranking over strategies is derived from the solution concept of obvious dominance, only OSP mechanisms are simple. Any mechanism that is not OSP is complex. If we instead assume that agents will never play weakly dominated strategies (in the usual sense), the set of simple mechanisms includes all strategy-proof mechanisms. More generally, as agents get more sophisticated, that is, as the partial order $\succ_{\theta_i}$ becomes more complete, the class of simple mechanisms expands. If agents can always totally rank all their strategies, all mechanisms are simple.

As is customary in mechanism design, we let the designer select which strategy player i should choose in case of indifference between S_i and S_i'.[12] Formally, we will treat $U_i(\theta_i)$ as consisting of equivalence classes induced by the indifference relation $\sim_{\theta_i}$, and we let the designer pick the representative of each equivalence class (we leave this description verbal in order not to further complicate our notation).

Strong and weak dominance of mechanisms. An advantage of a simple mechanism from the point of view of the designer is that she can predict how players will behave. In contrast, if any player is strategically confused, the designer—based only on Assumption 1—cannot determine the path of play. This seems to provide a strong argument in favor of simple mechanisms. However, that benefit is diminished if the designer can achieve better outcomes using a mechanism that confuses some types.

To formalize this idea, fixing a game Γ, let $\boldsymbol{S}_i$ denote a type-strategy for player i, that is, $\boldsymbol{S}_i(\theta_i)$ is the strategy selected by type θ_i of player i. We let $\boldsymbol{S}_i \subset U_i$ mean that $\boldsymbol{S}_i$ is a selection from the correspondence of undominated strategies, i.e., $\boldsymbol{S}_i(\theta_i) \in U_i(\theta_i)$ for all $\theta_i \in \Theta_i$. Let g be the outcome function mapping a profile of strategies $S = (S_1, ..., S_N)$ into an outcome in $\mathcal{X}$. Abusing notation slightly, let $v(\boldsymbol{S}) = \mathbb{E}_{\theta\sim\pi}\left[v(g(\boldsymbol{S}_1(\theta_1), ..., \boldsymbol{S}_N(\theta_N)), \theta)\right]$ denote the expected payoff of the designer when each player i behaves according to $\boldsymbol{S}_i$. We define a correspondence

$$V(\Gamma) = \text{Conv. Hull}\left(\{v(\boldsymbol{S}) : \{\boldsymbol{S}_i \subset U_i\}_{i\in\mathcal{N}}\}\right)$$

[12] One justification for this assumption is that the mechanism is paired with "recommended" strategies for the players, and the players follow the recommendation when they are indifferent.

to be the range of the designer's expected payoffs over all possible ways in which confused types can resolve their strategic confusion.[13] By definition, $V(\Gamma)$ is a singleton when Γ is a simple mechanism but it may be a (closed) interval when Γ is complex.

Definition 3 (Strong dominance)**.** *A mechanism Γ is strongly dominated if there exists a mechanism Γ' such that*

$$\max V(\Gamma) < \min V(\Gamma').$$

Intuitively, a mechanism Γ is strongly dominated by a mechanism Γ' if the expected payoff for the designer under Γ' is strictly larger than the expected payoff under Γ, regardless of how players choose from undominated strategies. Throughout, we will primarily apply this definition to the case in which Γ is the best simple mechanism (that is, Γ maximizes the designer's expected payoff among all simple mechanisms), and hence can only be strongly dominated by a complex mechanism.

In settings in which the value of simplicity is instrumental, that is, when the designer only cares about her ultimate expected utility, there is no reason for the designer to choose a simple mechanism over a complex mechanism that strongly dominates it. Later, we will also study a weaker version of that criterion, called weak dominance, under which the complex mechanism is guaranteed to yield weakly higher expected utility to the designer and may sometimes yield strictly higher expected utility (compared to the best simple mechanism).

2.1 Discussion

Modeling the solution concept via an arbitrary partial order on strategies paired with the behavioral assumption (Assumption 1) gives us significant flexibility but requires careful interpretation. We illustrate our definitions with a few examples.

The most straightforward application of our framework is to solution concepts in which the partial order $\succ_{\theta_i}$ depends solely on player i's payoffs and the structure of the game, but not on any characteristics of the other players. Apart from the already discussed obvious dominance, the classical notion of weak dominance, the strong obvious dominance, and all intermediate solution concepts discussed in Pycia and Troyan (2023) can be captured by our model. When the partial order is derived from weak dominance,

[13] Note that when computing the range of designer's payoffs, we assume that which undominated strategy a player selects cannot depend on the realization of other players' types.

simple mechanisms are exactly dominant-strategy mechanisms. When the partial order captures strong obvious dominance, simple mechanisms correspond to the class of strong obviously strategy-proof mechanisms of Pycia and Troyan (2023).

The partial order $\succ_{\theta_i}$ may also depend on the characteristics of other players and their assumed level of rationality. For example, we can define $\succ_{\theta_i}$ in such a way that $U_i(\theta_i)$ includes all strategies that survive K iterations of elimination of strictly dominated strategies. The level of rationality could be player-specific, allowing the modeler to capture differences in sophistication across players. "Stategically simple" mechanisms of Börgers and Li (2019) correspond to simple mechanisms in our terminology when the partial order is derived by assuming that agents have correct first-order beliefs about their opponents' preferences, and assume that their opponents are rational.

Finally, the flexibility of the partial order allows us to capture a number of behavioral phenomena that are not strategic in nature. For example, the partial order could reflect computational limitations (two strategies are not comparable if it is computationally infeasible to check which one performs better), context-dependent choice (the ranking of two strategies may depend on what other strategies are available to the player), or framing (two strategically equivalent games have different partial orders). The meaning of simplicity has to be adjusted accordingly. For instance, under computational limitations, a mechanism is simple if all agents can verify optimality of their prescribed strategy in a computationally feasible way.

Participation constraint. Even when simple mechanisms are strongly dominated, a possible argument in their favor is that players may be discouraged from participation if they face a complex mechanism. Moreover, in many settings (including the ones in which the designer's objective is to maximize her revenue), it is necessary to impose a participation constraint for the problem to be well defined.

To model participation, we may assume that all mechanisms $\Gamma \in \mathcal{M}$ contain a special strategy—called non-participation and denoted $S_i^{\emptyset}$—that is available to every player. For a simple mechanism, it is natural to model voluntary participation by requiring that the unique undominated strategy must dominate the non-participation strategy. We introduce two extensions of this condition to complex mechanisms. We say that a mechanism Γ provides *partial incentives to participate* if for all $i \in \mathcal{N}$, all $\theta_i \in \Theta_i$, there *exists* an undominated strategy $S_i \in U_i(\theta_i)$ that dominates $S_i^{\emptyset}$ (equivalently,

$S_i^{\emptyset} \notin U_i(\theta_i)$). We say that a mechanism Γ provides *full incentives to participate* if for all $i \in \mathcal{N}$, all $\theta_i \in \Theta_i$, *all undominated strategies* $S_i \in U_i(\theta_i)$ dominate $S_i^{\emptyset}$. The notion of partial incentives to participate is more appropriate when the participation decision is made at the outset of the game. The condition states that the player has a strategy in the ensuing game that she can identify as superior to non-participation. The notion of full incentives to participate is more appropriate if non-participation is thought of as an option for each player to walk away from the mechanism at any point, including after deciding which strategy to play.

Randomized mechanisms. In some cases, the designer may want to rely on randomization in the mechanism. Our modeling assumptions allow us to capture randomization as a special case of the framework. Indeed, one can think of player N in the game as "Nature" endowed with a type space and a probability distribution over types chosen by the designer (independent of other player's types), as well as a partial order that features indifference between all strategies in any mechanism. Then, the designer can choose any distribution over Nature's actions in the game, which replicates the effects of having explicit randomization at some nodes of the game tree. Of course, how other players reason about Nature's moves is then captured by the solution concept τ. For example, under the solution concept of obvious dominance, it is customary to treat Nature just like any other strategic player: When evaluating the worst-case payoff from a strategy S_i, player i takes into account all possible realizations of randomization performed by the designer (this is what we assume when defining obvious dominance in Appendix A.1).

Mixed strategies. We assumed that the set of strategies available to any player i in a game $\Gamma \in \mathcal{M}$ is finite, which guaranteed that for each strategy that is dominated, there exists another strategy that dominates it but is itself undominated. We interpreted $\mathcal{S}_i$ as the set of pure strategies available to player i. Consequently, the implicit assumption is that the solution concept τ only pins down agents' preferences over pure strategies. We can accommodate mixed strategies for solution concepts where for each strategy (pure or mixed) that is dominated, there exists a pure strategy that dominates it but is undominated according to the partial order.

3 Main results

In this section, we study strong dominance of simple mechanisms. Section 3.1 identifies environments in which the best simple mechanism may be strongly dominated. Section 3.2 provides a condition under which the best simple mechanism is not strongly dominated.

3.1 Simple mechanisms may be overly restrictive

A weakness of simple mechanisms is that they can be overly restrictive by requiring that *no* type should be confused. In some settings, the presence of certain preference types implies that the set of outcome functions implementable by simple mechanisms is small. However, the behavior of some agents' types could be insignificant for the designer, either because these types do not contribute to the designer's payoff, or they have low probability. In such cases, the designer may want to impose simplicity only for a subset of types with the remaining types being potentially confused.

To formalize this idea, we introduce a property that we call the "accommodation of additional types" (AAT). We show that if the AAT property is violated, the best simple mechanism is strongly dominated by a complex mechanism for at least some preferences of the designer. By means of an example (application of the concept of obvious dominance to a voting game), we show that failure of AAT may lead to strong dominance of the best simple mechanism under a natural objective function for the designer.

For a subset $\Theta' \subset \Theta$ of the type space, we say that a mechanism Γ is simple on Θ', if no type $\theta_i \in \Theta'_i$ of any player i is strategically confused. We denote by $\Theta \setminus \{\theta_i\}$ the type space $\Theta_1 \times \ldots \times (\Theta_i \setminus \{\theta_i\}) \times \ldots \times \Theta_N$.

Definition 4. *Fixing the primitive environment, we say that the accommodation of additional types (AAT) property holds if for any i, θ_i, and any mechanism $\Gamma' \in \mathcal{M}$ that is simple on $\Theta \setminus \{\theta_i\}$, there exists a mechanism $\Gamma \in \mathcal{M}$ that is simple on Θ and implements the same outcome as Γ' on $\Theta \setminus \{\theta_i\}$.*

The AAT property says that for any simple mechanism on $\Theta \setminus \{\theta_i\}$ we can always "accommodate" an additional type θ_i of agent i, that is, assign a dominant strategy to θ_i while keeping the outcome of the mechanism for the remaining types unchanged. To gain some intuition, fix some mechanism Γ' that is simple on $\Theta \setminus \{\theta_i\}$. To find the required mechanism Γ that is simple also for type θ_i, the designer has three possibilities. First, if type θ_i has a dominant strategy among the strategies already offered by Γ', then setting $\Gamma = \Gamma'$ satisfies the definition. This implies that the AAT property holds for solution concepts that induce a total order $\succeq_{\theta_i}$; it also holds for many classical solution concepts (like dominant-strategy implementation) in environments satisfying an appropriate single-crossing condition. Second, the designer can accommodate the additional type θ_i by adding a new strategy S_i to the game Γ' that is dominant for θ_i but dominated for all types $\theta_i' \neq \theta_i$. However, one must verify that the addition of the new strategy S_i does not alter the set of undominated strategies for other players. For example, under the solution concept of obvious dominance, the AAT property always holds in settings with a single agent for deterministic mechanisms (but may fail when mechanisms are stochastic—see Subsection 3.1.1). Finally, it may be possible to "redesign" the mechanism Γ' more substantially in a way that accommodates the type θ_i while preserving the outcomes for other types. When it is not possible to accommodate additional types, strong dominance of the best simple mechanism may occur.

Proposition 1. *Suppose that the AAT property fails. Then, there exists a utility function for the designer such that the best simple mechanism is strongly dominated.*

The intuition behind Proposition 1 is straightforward. Failure of the AAT property implies that the designer faces a trade-off: Simplicity of the mechanism comes at the cost of excluding the possibility of implementing certain outcomes conditional on a subset of types Θ'. If the designer's objective function is invariant to the outcomes implemented conditional on the remaining types $\Theta \setminus \Theta'$, then the trade-off is resolved in favor of implementing the most preferred outcome on Θ' and letting the types in $\Theta \setminus \Theta'$ be strategically confused. Effectively, if some types are problematic for the designer in the sense that their presence restricts the set of simple mechanisms, the designer may be better off ignoring these types when designing the mechanism. The proof of Proposition 1—which formalizes these arguments—can be found in Appendix A.3. It is easy to adapt the proof to show a version of Proposition 1 in which the designer's objective is fixed, but the probability of the "problematic types" becomes small enough.[14]

Next, we illustrate Proposition 1 with an example in which the partial order over stategies is generated by obvious dominance.

[14]See an earlier version of the paper Li ⓡ Dworczak (2022) for a formalization.

Example 3. Consider a voting environment with two agents and three alternatives, $\mathcal{X} = \{a, b, c\}$. Each agent's type can be represented as a ranking of the three alternatives. More specifically, each agent gets utility 1 if her top choice is implemented, 1/2 if her second choice is implemented, and 0 otherwise. The distribution of types π is i.i.d. uniform. The designer would like to maximize welfare but is Rawlsian and risk-averse: If u_i is the ex-post utility of agent i, then the designer's payoff is $v(\min(u_1, u_2))$ for some strictly concave and increasing function v.

The best OSP mechanism is dictatorship with full range.[15] The outcome function of that mechanism (with the row player being a dictator) is illustrated in Table 1.

Table 1: The best simple mechanism

	abc	*acb*	*bac*	*bca*	*cab*	*cba*
abc	a	a	a	a	a	a
acb	a	a	a	a	a	a
bac	b	b	b	b	b	b
bca	b	b	b	b	b	b
cab	c	c	c	c	c	c
cba	c	c	c	c	c	c

Table 2: A complex mechanism

	abc	*acb*	*bac*	*bca*	*cab*	*cba*
abc	a	a	a	a	a	a
acb	a	a	a	a	a	a
bac	b	b	b	b	b	b
bca	b	b	b	b	b	b
cab	a	a	a	c	c	c
cba(1)	a	a	a	c	c	c
cba(2)	b	b	b	b	b	b

Consider instead a delegation mechanism in which the designer specifies a set of menus from which the row player chooses one, with the other player selecting an alternative from the chosen menu. The row player can choose menu $\{a\}$, menu $\{b\}$, or menu $\{a, c\}$.

Types *abc*, *acb*, *bac*, *bca* of the row player retain their (obviously) dominant strategies. Type *cab* has a dominant strategy to choose the menu $\{a, c\}$. However, type *cba* does not have a dominant strategy: Both menu $\{b\}$ and menu $\{a, c\}$ are not dominated—the mechanism is complex. This is illustrated in Table 2.

However, whichever strategy type *cba* chooses, the expected payoff to the designer is strictly higher than in the best simple mechanism. Indeed, consider first the case in which type *cba* chooses menu $\{a, c\}$. Then, the difference in expected payoffs to the designer between the complex mechanism and the best simple mechanism (which can be calculated by comparing the cells of the two tables with different outcomes) is $\frac{1}{36} \times [v\,(1/2) - v(0)] > 0$. Now, consider the case in which type *cba* chooses menu b. The

[15] Given the definition of obvious dominance (see Appendix A.1), it is without loss of optimality to look at deterministic mechanisms, and thus only ordinal preferences of players matter. We can then directly verify that dictatorship with full range is optimal for our objective function.

difference in the expected payoffs of the designer is $\frac{1}{36} \times [-2v(1) + 4v(1/2) - 2v(0)] > 0$, by strict concavity of v. Thus, the complex mechanism is guaranteed to yield a strictly higher expected payoff to the designer regardless of how type *cba* resolves her confusion.

Note that the AAT property is violated in the above implementation environment: The simple mechanism defined by the first 5 rows of Table 2 by excluding type *cba* cannot be extended to a simple mechanism with type *cba* added back. While type *cba* influences the designer's payoff, the designer nevertheless prefers to "ignore" that type by implementing a mechanism that is simple for the remaining types and that makes *cba* strategically confused. $\square$

More generally, the AAT property fails in many classical social choice environments. It is well known that the only strategy-proof mechanisms whose range contains at least three alternatives are dictatorships; however, there are nontrivial strategy-proof social choice functions on restricted domains. Indeed, much of the research on strategy-proof social choice can be seen as establishing possibility results for (various) restricted domains.[16] Thus, the AAT property fails. Proposition 1 then implies that in the social choice environment, it might be beneficial to "ignore" some types and employ a social choice rule that is implementable on a smaller domain, rather than using a strategy-proof mechanism on a larger domain. In particular, this will hold if the "problematic" types occur with sufficiently low probability. For example, it is known that a social choice function on profiles of single-peaked preferences over a totally ordered set is strategy-proof if and only if it is a generalized median voter scheme. If the designer finds the outcome function of some generalized median voter scheme more desirable than that of a dictatorship, and the probability that the true type profile is contained in the set of single-peaked preferences is high enough, then the generalized median voter scheme strongly dominates the best simple mechanism on the full domain. Similar examples can be found for the solution concept of OSP; for example, Arribillaga et al. (2020) and Bade and Gonczarowski (2017) characterize the class of OSP-implementable and unanimous social choice functions for single-peaked preferences.

3.1.1 Confusing a single agent via randomization

Consider the case of a single agent, and suppose that the agent has complete preferences over outcomes in $\mathcal{X}$. Then, the AAT property trivially holds if we restrict attention

[16]We refer interested readers to Barberà (2010) for a survey on strategy-proof social choice.

to deterministic mechanisms (and assume that the agent can predict the outcome of every strategy).[17] Indeed, an additional type of the agent can always be accommodated by letting that type choose among strategies already made available to other types. In fact, AAT must hold by Proposition 1 because simple mechanisms cannot be strongly dominated when the partial order $\succeq_{\theta_1}$ is complete.

If we allow for randomization in the mechanism (and the agent is not Bayesian), then the AAT property may fail. This is not surprising: As we explained in Section 2, we capture randomization in the mechanism by introducing an auxiliary player (Nature), so the mechanism effectively has two players. However, Proposition 1 is unsatisfactory in this case because it allows the designer's objective function to depend on the outcome of randomization (Nature's type). Thus, Proposition 1 does not resolve the question of whether the best simple mechanism can become strongly dominated by confusing the agent with randomization.

In Appendix A.4, we answer this question affirmatively by constructing an example of strong dominance in the single-agent setting under the solution concept of OSP. The example is based on the observation that the designer can use a randomization device to implement an outcome that is not implementable as ex-ante randomization over simple mechanisms. It is known (see, for example, Ashlagi and Gonczarowski (2018) and Pycia and Troyan (2023)) that randomization cannot increase the designer's payoff within the class of OSP mechanisms. Interestingly, the example shows that this does not imply that the designer should never randomize when facing unsophisticated agents. On the contrary, randomization can sometimes be used to purposefully confuse the agents in order to obtain a superior outcome.

3.2 An optimality foundation for simple mechanisms

In this subsection, we provide conditions under which the best simple mechanism is not strongly dominated. The key observation is that for any finite mechanism Γ (simple or complex), $\min V(\Gamma)$ is weakly less than the maximum expected payoff the designer could obtain in a mechanism that only satisfies a subset of incentive constraints that correspond to the edges of an arbitrary tree in the type space. Thus, if the designer's expected payoff from the best simple mechanism is the same as that in the relaxed

[17] The same conclusions holds if the agent is Bayesian, and $\mathcal{X}$ already contains all randomized outcomes.

problem with incentive constraints that correspond to the edges of some tree in the type space, then the best simple mechanism is not strongly dominated. This argument builds on the insights of Yamashita (2015) developed for implementation in weakly undominated strategies.

Let $G = (V, E)$ be a directed graph with vertex set V and edge set $E \subseteq V \times V$. A graph is called a (rooted) tree if it has exactly one vertex with no outgoing edges (called the "root") and exactly one path from every vertex to the root. For each agent i, consider a tree $T_i = (\Theta_i, E_i)$ where $E_i \subseteq \Theta_i \times \Theta_i$. Each directed edge $(\theta_i, \theta_i') \in E$, also denoted $\theta_i \to \theta_i'$, corresponds to the incentive constraint that type θ_i does not want to adopt the strategy of type θ_i'. For each agent i, fix a tree T_i. The collection of trees $\{T_i\}_{i \in \mathcal{N}}$ then defines a relaxed optimization problem in which the only incentive constraints are the ones that correspond to the edges on the trees. Formally, a mechanism $\Gamma \in \mathcal{M}$ is *feasible* for the IC-relaxed problem if there exists an assignment of strategies $\boldsymbol{S}_i(\theta_i)$ to each type θ_i of each player i such that $\boldsymbol{S}_i(\theta_i) \succ_{\theta_i} \boldsymbol{S}_i(\theta_i')$ whenever $\theta_i \to \theta_i'$. Clearly, this is a weaker requirement than forcing the mechanism Γ to be simple (which would correspond to imposing the condition for the complete graph on the type space Θ_i). The designer's payoff from the IC-relaxed optimization problem is then defined as the supremum over expected payoffs for the designer generated by feasible mechanisms evaluated at any assignment of strategies that makes them feasible.

Proposition 2. *Suppose that there exists a collection of trees $\{T_i\}_{i \in \mathcal{N}}$ such that the designer's expected payoff from the IC-relaxed optimization problem corresponding to $\{T_i\}_{i \in \mathcal{N}}$ is the same as that from using the optimal simple mechanism. Then, the optimal simple mechanism is not strongly dominated.*

We now explain how to incorporate participation constraints. Recall from Section 2 that we model participation decisions by assuming that each feasible mechanism must include a non-participation strategy $S_i^{\emptyset}$ for every player. We define the IR-non-relaxed problem by requiring that every type of every player is assigned a strategy that dominates non-participation. We define the IR-relaxed problem by only requiring that the strategy assigned to the root of each tree T_i dominates non-participation.

Proposition 2'. *Suppose that there exists a collection of trees $\{T_i\}_{i \in \mathcal{N}}$ such that the designer's expected payoff from the (i) IR-relaxed and IC-relaxed or (ii) IR-non-relaxed and IC-relaxed optimization problem corresponding to $\{T_i\}_{i \in \mathcal{N}}$ is the same as that from* *using the optimal simple mechanism with participation constraints. Then, the optimal simple mechanism is not strongly dominated by any mechanism with, respectively, (i) partial or (ii) full incentives to participate.*

Proposition 2(2'), while abstract, offers a concrete procedure to establish optimality foundations for simple mechanisms. For illustration, we revisit the two examples in the introduction. Applying Proposition 2', we show that the best simple mechanism is not strongly dominated by any mechanism with full incentives to participate in Example 1, and the best simple mechanism is not strongly dominated by any mechanism with partial (and hence full) incentives to participate in Example 2.

Example 4 (Example 1 revisited)**.** Recall that the best simple mechanism gives the designer an expected profit of $1/5$. It can be directly verified (by considering all possible trees for both players, see Appendix A.2) that the trees that yield the lowest value of the relaxed problem are $T_A = \{\{0,\, 2/3\},\, 0 \to 2/3\}$ and $T_B = \{\{1/3,\, 1\},\, 1 \to 1/3\}$. That is, for trader A we only impose the IC constraint that type 0 does not want to imitate type $2/3$, and for trader B that type 1 does not want to imitate type $1/3$. The value of the IC-relaxed and IR-non-relaxed problem is $1/5$. Thus, the best simple mechanism is *not* strongly dominated by any mechanism with *full* incentives to participate, by Proposition 2'.

The superior complex mechanism (which satisfies the partial incentives to participate) from Example 1 can be understood as partially capturing the benefits of the relaxed IR constraint of type 0 of trader A. Indeed, if we removed the first offer to trader A from the mechanism, type $2/3$'s optimal choice would be unaffected (since accepting the first offer was obviously dominated for that type anyway), while type 0 would no longer be confused—she would follow the same strategy as type $2/3$. However, type 0 would lose its partial incentives to participate since following $2/3$'s strategy sometimes results in a strictly negative payoff. This reveals that the reason for making the first offer to trader A in the complex mechanism is to ensure participation. By confusing type 0 of trader A, the designer extracts more surplus with one undominated strategy ("reject the first offer but accept the next offer") while dominating non-participation with a second undominated strategy ("accept the first offer"). $\square$

Failure of the premise of Proposition 2 or 2' indicates that a strict improvement could be made in the problem of finding the best simple mechanism if some IC or IR constraint were relaxed. In a simple mechanism, however, a strategy assigned to each type must satisfy all these constraints (that is, no IC or IR constraint can be relaxed), and hence it is not possible to achieve that potential improvement. In a complex mechanism, by contrast, each type may be assigned multiple undominated strategies (as in Example 1 for the case of partial incentives to participate) and, at least in principle, no single strategy must satisfy all the constraints. This additional flexibility of complex mechanism may sometimes allow the designer to achieve the improvement.

Example 5 (Example 2 revisited)**.** Proposition 2', adopted to the solution concept of SP, can be used to show that in a large class of environments with a revenue-maximizing designer—including single-unit auctions—the optimal SP is not strongly dominated by any mechanism with partial incentive to participate.[18] Thus, if the optimal SP mechanism can be implemented via an OSP mechanism, as in Example 2, the optimal OSP mechanism is not strongly dominated by any mechanism with partial incentives to participate (and hence not strongly dominated by any mechanism with full incentives to participate).

4 Weak dominance

In this section, we briefly investigate the concept of *weak dominance*.

Definition 5 (Weak dominance)**.** *A mechanism* Γ *is weakly dominated if there exists a mechanism* Γ' *such that*

$$\max V(\Gamma) \leq \min V(\Gamma') \text{ and } \max V(\Gamma) < \max V(\Gamma').$$

Weak dominance of the best simple mechanism by a complex mechanism gives the designer a guarantee of a weakly better payoff together with the *possibility* of a strictly better payoff. This is a less demanding requirement (compared to strong dominance) for a complex mechanism to be deemed superior. Nevertheless, it highlights a sense in

[18]Formally, this holds when the uniform shortest-path tree condition holds and the distribution π is regular, as formulated by Chen and Li (2018). This covers many classical applications of mechanism design such as single-unit auctions (e.g., Myerson (1981)), public goods (e.g., Mailath and Postlewaite (1990)), standard bilateral trade (e.g., Myerson and Satterthwaite (1983)) and also some multi-dimensional environments such as the auction for capacity constrained bidders (see Malakhov and Vohra (2009)). The detailed analysis can be found in Li ⓡ Dworczak (2022).

which the designer cannot lose (and has the potential to gain) by opting for the complex mechanism.

By means of an extended example—the problem of maximizing revenue over OSP mechanisms—we argue that weak dominance of simple mechanisms by complex ones is a common occurrence. We also provide one positive result: Under additional conditions, the posted price mechanism is not weakly dominated when agents are assumed to avoid obviously dominated strategies.[19]

We analyze a standard revenue-maximization problem with quasi-linear utilities. Let X be the space of possible allocations (which could involve randomization), and define $\mathcal{X} = X \times \mathbb{R}^N$, with $(y, t_1, ..., t_N) \in \mathcal{X}$ interpreted as an outcome in which allocation y is implemented, and player i pays the designer t_i. We have $u_i((y, t_1, ..., t_N), \theta_i) = \tilde{u}_i(y, \theta_i) - t_i$, for some arbitrary $\tilde{u}_i(y, \theta_i)$ assumed non-negative and non-constant in θ_i. The designer maximizes expected revenue. We will consider two cases. In the first case, the designer is satisfied with a mechanism that provides partial incentives to participate; we will show that this leads to the best simple mechanism being weakly dominated in a particularly blatant way, with revenue that is unbounded in the best case for the designer. In the second case, the superior mechanism will provide full incentives to participate. In both cases, we normalize the payoff from the non-participation strategy to be zero.

Throughout, we assume that the partial order on strategies is derived from obvious dominance, as defined in Appendix A.1. The proofs of all claims can be found in Appendix A.6.[20]

Claim 1. *The revenue-maximizing OSP mechanism is weakly dominated by a complex mechanism with partial incentives to participate.*

The superior complex mechanism exploits—in a very stark way—the possibility that agents lack strategic sophistication. The designer approaches some agent i and proposes to her the following additional bet: Agent i gets a large amount of money M if a coin is heads but pays the designer M if the coin is tails. The designer biases the coin so that tails has probability $1 - \epsilon$ for an arbitrarily small ϵ. The agent finds

[19] In an independent paper, Mukherjee et al. (2024) provide sufficient conditions for a social choice correspondence to be implemented in weakly undominated strategies. They apply their results to several economic environments; their results are broadly consistent with and complementary to our analysis on weak dominance. The message of this section is reinforced by additional examples of weak dominance described in Mukherjee et al. (2024).

[20] The proofs are based on a more general result that can be used in other contexts to identify weakly-dominating mechanisms.

accepting the bet OSP-undominated because she evaluates her outcomes conditional on each realization of the randomization device: When M is sufficiently large, taking the bet dominates all other strategies conditional on the coin being heads. If the agent refuses to take the bet, the designer is still guaranteed the same revenue as in the original simple mechanism; and if the agent accepts the bet, the designer's expected revenue is unbounded.

Reinforcing the message of Subsection 3.1.1, we conclude that—even though randomization cannot increase the designer's expected payoff from the best OSP mechanism—it might nevertheless be beneficial to use against unsophisticated agents.

Even if agents were assumed to be Bayesian with respect to the designer's randomization, the designer could still achieve an unbounded revenue in the best case of a complex mechanism when $N \geq 2$. The additional option for agent i has the same two outcomes for i, but the choice between them is made by some other agent j who is paid $\epsilon > 0$ to choose the outcome in which agent i pays M. Accepting this modified "bet" is OSP-undominated for i because agent i is not assumed to believe that agent j will never play a dominated strategy.

The above weakly-dominating mechanisms may seem unlikely to "work" in practice, in the sense that the additional strategic option offered in the complex mechanism is clearly "unattractive" for the agent. We offer two comments: First, similarly to how many results are interpreted in mechanism design, we view the value of these examples as illustrating the possibility that a simple mechanism may be dominated. Our construction is optimized for making the mathematical argument concise; there may exist more subtle ways to weakly dominate a simple mechanism. Second, anecdotally, sellers frequently offer seemingly unattractive options to customers hoping to exploit their potential inability to rank these options as inferior.[21] And in any case, the designer never loses by switching to a superior complex mechanism, so she may prefer the complex mechanism even if she thinks that the agent is unlikely to choose the additional option.

A different criticism of the above complex mechanisms is that the agent will sometimes walk away with a negative payoff. We thus turn attention to the case when the complex mechanism is required to provide full incentives to participate. To simplify exposition, we consider the classical problem of allocating a single object to one of N

[21] See, for example, Chernev et al. (2015) for a survey of results on choice overload.

ex-ante identical bidders.[22]

Claim 2. *Suppose that $N \geq 2$, and let $\bar{u}$ be the highest possible valuation for the object. If it is not a revenue-maximizing OSP mechanism to sequentially offer the object at a price of $\bar{u}$ to all the players, then the best OSP mechanism is weakly dominated by a complex mechanism with full incentives to participate.*

The proof resembles the mechanism we described previously for the case $N \geq 2$. The main difference is that the additional option is carefully constructed so that it always yields a non-negative payoff to the types that may choose it.

The assumption of Claim 2 cannot be completely relaxed. For a simple example, note that if all players have a value of $\bar{u}$ for the object with probability one, then sequentially offering the object at a price of $\bar{u}$ to all the players is an optimal simple mechanism that is *not* weakly dominated under full incentives to participate (for it to be dominated, there would have to exist an on-path history in which some player is charged more than $\bar{u}$ but that would be incompatible with full incentives to participate).

Next, we prove a general positive result for the case of $N = 1$.

Claim 3. *Suppose that the designer sells a single indivisible object to a single agent, attempting to maximize revenue. Assume that there exists a unique optimal OSP mechanism (in which case it must be outcome-equivalent to a posted price mechanism). Then, that mechanism is not weakly dominated by any complex mechanism with full incentives to participate.*

We emphasize that *not* being weakly dominated is a particularly strong optimality foundation; it implies that any complex mechanism that sometimes leads to a higher payoff for the designer must necessarily be strictly worse in some other case.

The proof of Claim 3 is relatively involved. The assumption that the optimal simple mechanism is unique (which holds generically) is needed. In Appendix A.6.5, we construct an example in which the optimal simple mechanism is not unique and we show that it is weakly dominated.

[22] Note that the ascending clock auction with jump-bidding in Example 2 in the introduction provides full incentives to participate under OSP. This is because if type $v \geq b$ makes a jump bid to b at price $p < b$ (jump-bidding is dominated for all other types), then she gets a non-negative payoff after every possible history. Obviously, her payoff is also non-negative if she does not jump-bid.

5 Conclusion

In mechanism design, it seems useful to distinguish (simple) mechanisms in which agents face a straightforward choice problem from (complex) mechanisms that require agents to engage in complex mental tasks to determine their optimal strategy. The literature has made a great deal of progress in terms of formulating different notions of simplicity and characterizing mechanisms that are simple according to these notions. However, the understanding of the design of mechanisms with unsophisticated agents, as we argued in this paper, is far from complete. Indeed, in many cases, the designer might prefer mechanisms that are not simple, even under the assumption that agents choose strategies that are the worst possible for the designer whenever they are confused.

We suggest some directions for further research. An important avenue is the optimal design of mechanisms when agents are strategically unsophisticated. Our analysis indicates that the optimal design of mechanisms with unsophisticated agents could be challenging: One should not simply optimize over the class of simple mechanisms, at least not without first establishing their optimality. Searching over all mechanisms presents new challenges, as the designer cannot rely on any standard revelation principle when mechanisms are evaluated by their worst-case performance.

Relatedly, while we focused primarily on negative results throughout the paper, we expect that establishing optimality foundations for simple mechanisms—not being weakly or strongly dominated—might be a particularly promising research direction. This is because such a foundation can often be found by first solving an easier relaxed problem, and then showing that the upper bound is achieved by a simple mechanism (see Section 3.2).

While we focused on obvious dominance as the leading application in this paper, our framework could be used to study a number of other solution concepts, even ones not traditionally associated with simplicity. Because we have not imposed any restrictions on the partial order on strategies that reflects agents' reasoning, the order could depend on agents' beliefs. This would allow us to study solution concepts such as Bayesian Nash equilibrium; depending on how the partial order is defined, simplicity would correspond to either robustness to selecting a best response at equilibrium, or robustness to equilibrium selection.[23] Relatedly, while we studied a private-value environment, questions of simplicity versus complexity are equally intriguing in the interdependent-value setting.[24]

Finally, it would be interesting to conduct experimental tests of the best simple mechanism against the complex mechanism that dominates it, such as the ascending-clock auction and the ascending-clock auction with jump bidding. These findings could then be used to support or invalidate the superiority of complex mechanisms.

References

Arribillaga, R. P., J. Massó, and A. Neme (2020): "On Obvious Strategy-Proofness and Single-Peakedness," *Journal of Economic Theory*, 186, 104992.

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. A. Gonczarowski (2017): "Gibbard-Satterthwaite Success Stories and Obvious Strategyproofness," in *Proceedings of the 2017 ACM Conference on Economics and Computation*, 565–565.

Barberà, S. (2010): "Strategy-Proof Social Choice," in *Handbook of Social Choice and Welfare*, ed. by K. Arrow, A. Sen, and K. Suzumura, North-Holland, vol. 2, 731–831.

Börgers, T. (1991): "Undominated Strategies and Coordination in Normalform Games," *Social Choice and Welfare*, 8, 65–78.

——— (2017): "(No) Foundations of Dominant-Strategy Mechanisms: A Comment on Chung and Ely (2007)," *Review of Economic Design*, 21, 73–82.

Börgers, T. and J. Li (2019): "Strategically Simple Mechanisms," *Econometrica*, 87, 2003–2035.

Börgers, T. and D. Smith (2012): "Robustly Ranking Mechanisms," *American Economic Review*, 102, 325–329.

[23] The first case can be obtained by fixing a BNE of the game, and then defining the partial order by evaluating the expected utility from every strategy against the opponents' equilibrium strategy profile; the second case can be obtained by letting all equilibrium strategies be undominated.

[24] In the interdependent-value setting, Jehiel et al. (2006) and Yamashita and Zhu (2022) consider a designer who does not have realizable information about the agents' beliefs, while the agents are still assumed to play a Bayesian equilibrium. While the focus of these papers differs from ours, we note that Example 5.1 in Jehiel et al. (2006) and the analysis in Yamashita and Zhu (2022) could be used to show that the designer might prefer a mechanism that is not ex post incentive compatible to the optimal ex post incentive-compatible mechanism in the interdependent-value setting, under the assumption that agents do not play weakly dominated strategies. Yamashita (2015) shows that for revenue maximization in an interdependent-value auction, under certain conditions, a version of a second-price auction (which is neither dominant-strategy nor ex post incentive compatible) is optimal for implementation in weakly undominated strategies.

Carroll, G. (2014): "A Complexity Result for Undominated-Strategy Implementation," *Working Paper*.

——— (2019): "Robustness in Mechanism Design and Contracting," *Annual Review of Economics*, 11, 139–166.

Chen, Y.-C. and J. Li (2018): "Revisiting the Foundations of Dominant-Strategy Mechanisms," *Journal of Economic Theory*, 178, 294–317.

Chernev, A., U. Böckenholt, and J. Goodman (2015): "Choice Overload: A Conceptual Review and Meta-Analysis," *Journal of Consumer Psychology*, 25, 333–358.

Cramton, P., R. Gibbons, and P. Klemperer (1987): "Dissolving a Partnership Efficiently," *Econometrica*, 55, 615–632.

Glazer, J. and A. Rubinstein (2012): "A Model of Persuasion with Boundedly Rational Agents," *Journal of Political Economy*, 120, 1057–1082.

——— (2014): "Complex Questionnaires," *Econometrica*, 82, 1529–1541.

Hartline, J. D. and T. Roughgarden (2009): "Simple versus Optimal Mechanisms," in *Proceedings of the 10th ACM conference on Electronic commerce*, 225–234.

Jackson, M. O. (1992): "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies*, 59, 757–775.

Jakobsen, A. M. (2020): "A Model of Complex Contracts," *American Economic Review*, 110, 1243–1273.

Jehiel, P., M. Meyer-ter Vehn, B. Moldovanu, and W. R. Zame (2006): "The Limits of Ex Post Implementation," *Econometrica*, 74, 585–610.

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

Li, J. ® P. Dworczak (2022): "Are Simple Mechanisms Optimal when Agents are Unsophisticated?" Available at SSRN: https://ssrn.com/abstract=3632971.

Loertscher, S. and L. M. Marx (2020): "A Dominant-Strategy Asset Market Mechanism," *Games and Economic Behavior*, 120, 1–15.

Lu, H. and J. Robert (2001): "Optimal Trading Mechanisms with Ex Ante Unidentified Traders," *Journal of Economic Theory*, 97, 50–80.

Mackenzie, A. (2019): "A Revelation Principle for Obviously Strategy-Proof Implementation," *Working Paper*.

Mailath, G. and A. Postlewaite (1990): "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies*, 57, 351–367.

Malakhov, A. and R. V. Vohra (2009): "An Optimal Auction for Capacity Constrained Bidders: A Network Perspective," *Economic Theory*, 39, 113–128.

Mukherjee, S., N. Muto, E. Ramaekers, and A. Sen (2019): "Implementation in Undominated Strategies by Bounded Mechanisms: The Pareto Correspondence and a Generalization," *Journal of Economic Theory*, 180, 229–243.

Mukherjee, S., N. Muto, and A. Sen (2024): "Implementation in Undominated Strategies with Applications to Auction Design, Public Good Provision and Matching," *Journal of Economic Theory*, 216, 105783.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Myerson, R. and M. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29, 265–281.

Pycia, M. and P. Troyan (2023): "A Theory of Simplicity in Games and Mechanism Design," *Econometrica*, 91, 1495–1526.

Ray, D. ® A. Robson (2018): "Certified Random: A New Order for Coauthorship," *American Economic Review*, 108, 489–520.

Spiegler, R. (2011): *Bounded Rationality and Industrial Organization*, Oxford University Press.

Yamashita, T. (2015): "Implementation in Weakly Undominated Strategies: Optimality of Second-Price Auction and Posted-Price Mechanism," *Review of Economic Studies*, 82, 1223–1246.

Yamashita, T. and S. Zhu (2022): "On the Foundations of Ex Post Incentive-Compatible Mechanisms," *American Economic Journal: Microeconomics*, 14, 494–514.

A Appendix

A.1 The partial order under the notion of OSP

We consider finite mechanisms that are imperfect-information, extensive-form games with perfect recall and consequences in $\mathcal{X}$. The designer may wish to use randomization in the mechanism. As explained in the main text, to incorporate this possibility, we can treat agent N as a dummy player ("Nature") with constant preferences. The distribution over Nature's strategies can then be picked by the designer. To shorten the exposition, we only introduce notation associated with a generic game Γ that we are going to use:

(1) $\mathcal{H}$ is the set of histories, with representative element h, and $h_\emptyset$ denoting the initial (empty) history;

(2) $\subset$ is the precedence relation over histories;

(3) $\mathcal{Z}$ is the set of terminal histories, with representative element z;

(4) $g(z) \in \mathcal{X}$ is the outcome resulting from z;

(5) I_i denotes an information set of agent i;

(6) $A(I_i)$ is the set of actions available at information set I_i;

(7) A (pure) strategy S_i chooses an action $a \in A(I_i)$ at every information set I_i of agent $i \in \mathcal{N}$, and $\mathcal{S}_i$ is the collection of all (pure) strategies for player i;

(8) A strategy profile $S = (S_1, \ldots, S_N)$ specifies a strategy for each player;

(9) $z(h, S)$ denotes the terminal history that results when we start at history h and play proceeds according to the strategy profile S.

We follow Li (2017) to define the partial order $\succ_{\theta_i}$ corresponding to the solution concept of obvious dominance. We say that the information set I_i is on the path of play of strategy S_i if there exists S_{-i} and $h \in I_i$ such that $h \subset z(h_\emptyset, S_i, S_{-i})$. Given two strategies S_i and S_i', we define $\beta(S_i, S_i')$ to be the set of information sets that are on the path of play of both S_i and S_i'. Under perfect recall, $I_i \in \beta(S_i, S_i')$ implies that S_i and S_i' choose the same actions at all information sets preceding I_i. If $I_i \in \beta(S_i, S_i')$ but S_i and S_i' choose different actions at I_i, then we call I_i an earliest point of departure for these two strategies. We let $\alpha(S_i, S_i')$ denote the set of all earliest points of departure for these two strategies.

Definition 6 (OSP). *$S_i \preceq_{\theta_i} S_i'$ if for all $I_i \in \alpha(S_i, S_i')$,*

$$\max_{h \in I_i, S_{-i}} u_i(g(z(h, S_i, S_{-i})), \theta_i) \leq \min_{h \in I_i, S_{-i}} u_i(g(z(h, S_i', S_{-i})), \theta_i).$$

We say that strategy S_i is obviously dominated by strategy S_i' for type θ_i of agent i if $S_i \prec_{\theta_i} S_i'$, that is, if $S_i \preceq_{\theta_i} S_i'$ but not $S_i' \preceq_{\theta_i} S_i$.

Thus, a strategy S_i is obviously dominated for type θ_i if there exists another strategy S_i' such that, starting at any earliest point of departure, the worst possible payoff under S_i' for type θ_i across all strategies of other players and Nature is higher than the best possible payoff under S_i for type θ_i across all strategies of other players and Nature. A strategy S_i is obviously dominant if all non-equivalent strategies S_i' are obviously dominated by it.

A.2 The optimal OSP mechanism for Example 1

It follows from Theorem 3.1 of Bade and Gonczarowski (2017) that it is without loss of optimality for OSP implementation to look at gradual revelation mechanisms (see Ashlagi and Gonczarowski, 2018, Pycia and Troyan, 2023 and Mackenzie, 2019 for related revelation principles for OSP mechanisms). In our simple example with two players and two types, this means that we can assume that in the best OSP mechanism, at the first decision node, one of the players ("leader") makes a binary decision (with the two types choosing different actions—potentially leading to the same outcome—as part of their obviously dominant strategies), and then in each of the two possible histories, having observed the choice of the leader, the other player ("follower") makes a binary decision.

Therefore, an upper bound on the profit in the optimal OSP mechanism can be derived in the following way. When the follower chooses her optimal action, she already knows the action chosen by the leader. OSP requires that for any choice of the leader, each type of the follower must weakly prefer her equilibrium strategy (action) to choosing the alternative action. Thus, each type of the follower must have a standard *dominant strategy* in the normal-form representation of the game. In contrast, when the leader chooses her action at the initial decision node, it must be that the worst possible payoff from choosing her equilibrium action (over the two possible actions that can be selected by the follower in the subgame) is weakly higher than the best possible payoff from choosing the alternative action. In the normal-form representation of the game, this can be captured by requiring that the payoff from the equilibrium strategy of each type of the leader under any choice of the strategy S_f for the follower is weakly higher than the payoff from following the alternative strategy under any choice of the strategy S'_f (where, importantly, S_f could be different from S'_f).

Summarizing, an upper bound can be derived by using a normal-form game with the usual strategy-proof constraints, except that for the leader the constraints are strengthened in the way described above. Crucially, all these constraints are linear in the allocation and transfers, and so is the objective function of the designer (intuitively, we avoid taking the min and max in the definition of obvious dominance by iterating over all possible *pairs* of strategies $(S_f,\, S'_f)$ for the follower when comparing the two strategies of the leader). Thus, we obtain a linear program that can be solved using standard linear-programming tools, yielding an upper bound of 1/5. This upper bound is achieved by the OSP mechanism described in the example, proving its optimality.

A.3 Proof of Proposition 1

If the AAT property fails, then there exists a type θ_i of some agent i and a mechanism Γ' that is simple on $\Theta \setminus \{\theta_i\}$ but cannot be extended to a simple mechanism on Θ with the same outcome on $\Theta \setminus \{\theta_i\}$. Because the mechanism Γ' is simple on $\Theta \setminus \{\theta_i\}$, it leads to a unique outcome conditional on these types that we can denote by a function $\lambda : \Theta \setminus \{\theta_i\} \to \mathcal{X}$. Define an objective function v of the designer by

$$v(x,\, \theta) = \mathbf{1}_{\{x=\lambda(\theta),\, \theta\in\Theta\setminus\{\theta_i\}\}}.$$

With the objective function v, the best simple mechanism on Θ must yield an optimal payoff strictly lower than $\pi(\Theta \setminus \{\theta_i\})$ because, by definition, the outcome λ cannot be implemented by a simple mechanism on Θ. Consider, however, the mechanism Γ'. This mechanism must be complex on Θ. However, the worst-possible expected payoff from using this mechanism is $\pi(\Theta \setminus \{\theta_i\})$ because conditional on $\Theta \setminus \{\theta_i\}$ the outcome λ is implemented, while the outcome implemented conditional on player i having type θ_i does not effect the designer's objective. Therefore, the optimal simple mechanism is strongly dominated.

A.4 Example for Subsection 3.1.1

We explicitly construct an example in which the best OSP mechanism is strongly dominated. There is one player with three equally-likely types, $\Theta = \{u, m, d\}$, with the following ordinal preferences over $\mathcal{X} = \{U, U', M, M', D, D'\}$:

1. type u: $M > U > D' > D > M' > U'$;
2. type m: $D > M > U' > U > D' > M'$;
3. type d: $U > D > M' > M > U' > D'$.

The designer gets a utility of 1 if the type is $j \in \{u, m, d\}$ and she implements outcome J or J'; she gets -1 otherwise.

Lemma 1. *The best simple mechanism is to implement any fixed (possibly random) outcome; the expected payoff for the designer is* $-1/3$.

Recall that it is without loss of optimality for the designer to consider deterministic mechanisms when optimizing in the class of simple (OSP) mechanisms. A deterministic simple mechanism for a single agent can be represented as a direct assignment of alternatives to types such that no type strictly prefers another type's assignment to her own. Suppose that there exists such an assignment that gives the designer an expected payoff strictly above $-1/3$. Then, at least two types $j \in \{u, m, d\}$ must be assigned either J or J'. If any type j is assigned J', then no other alternative can be offered by the mechanism since j ranks J' last. Thus, the mechanism must offer J to two distinct types j. However, that's a contradiction because at least one of these types would prefer the allocation of the other one, no matter which two types $j \in \{u, m, d\}$ we choose.

To finish the proof, we construct a superior complex mechanism Γ.

Lemma 2. *There exists a complex mechanism* Γ *that guarantees the designer an expected payoff of* 0.

In the mechanism Γ, the designer uses a randomization device that has equally likely outcomes H and T, and offers three possible strategies to the agent, as represented by the following normal form:

	H	T
S_U	U	D'
S_M	M	U'
S_D	D	M'

Type $j \in \{u, m, d\}$ is confused between the two strategies offering J and J', respectively, with the former one (denoted S_J) leading to the 2nd or 3rd alternative, and the latter to the 1st or 6th alternative. However, the remaining strategy is obviously dominated for j by the strategy S_J as it leads to the 4th or 5th alternative. Regardless of how j resolves her strategic confusion, either J or J' is implemented with probability $1/2$, and thus the designer obtains 0 in expectation.

A.5 Proof of Proposition 2 and 2'

We prove Proposition 2 first, and then explain how to modify the steps to obtain Proposition 2'. We start with a lemma that builds on the insight in Yamashita (2015, Theorem 1), which can be viewed as its generalization to a larger class of environments and our abstract notion of a solution concept.

Lemma 3. *For any mechanism Γ, $\min V(\Gamma)$ is upper-bounded by the value of the IC-relaxed problem corresponding to any fixed collection of trees $\{T_i\}_{i\in\mathcal{N}}$ (as defined in Section 3.2).*

Proof. Fix an arbitrary finite mechanism Γ, an agent i, and a tree T_i. Let $T_i^+(\theta_i) = \{\theta_i' : \theta_i' \to \theta_i\}$ be the set of types who point towards type θ_i in the tree T_i. Consider the following procedure. Starting at the root of the tree T_i—which is some type θ_i^0 with no edges coming out of it—select any undominated strategy for θ_i^0, $S_i^0 \in U_i(\theta_i^0)$. Next, for any type $\theta_i' \in T_i^+(\theta_i^0)$, we can find an undominated strategy $S_i' \in U_i(\theta_i')$ that either dominates S_i^0 or is equal to S_i^0 (this step uses finiteness of the mechanism; if S_i^0 is not in $U_i(\theta_i')$, then there must exist a strategy in $U_i(\theta_i')$ that dominates it). We proceed inductively. Once some type θ_i is assigned a strategy, we assign undominated strategies to all types $T_i^+(\theta_i)$ that either equal or dominate the strategy assigned to θ_i. Because T_i is a finite tree, this procedure must stop at some point, with every type being assigned a strategy. The same procedure is carried out for all other agents.

Because each type is assigned a strategy from $U_i(\cdot)$ in the procedure, when all types execute their assigned strategies, the expected payoff $\bar{v}$ to the designer must weakly exceed $\min V(\Gamma)$ (which is the outcome of the designer-adversarial selection from $U_i(\cdot)$). Moreover, the procedure guarantees that the mechanism—along with the assignment of strategies—is feasible for the IC-relaxed problem corresponding to the collection of trees $\{T_i\}_{i\in\mathcal{N}}$. Therefore, $\bar{v}$, and hence also $\min V(\Gamma)$, is weakly below the value of the IC-relaxed problem. $\square$

Proposition 2 follows immediately: By assumption, there exists a collection of trees $\{T_i\}_{i\in\mathcal{N}}$ such that the value of the IC-relaxed problem corresponding to $\{T_i\}_{i\in\mathcal{N}}$ is the same as the designer's expected payoff from the best simple mechanism. Hence, by applying Lemma 3 for the collection $\{T_i\}_{i\in\mathcal{N}}$, we conclude that there cannot exist a mechanism Γ with $\min V(\Gamma)$ strictly above the expected payoff of the best simple mechanism, and hence the best simple mechanism is not strongly dominated.

We now explain how to incorporate participation constraints in the above procedure to obtain Proposition 2'. Suppose that the superior complex mechanism is required to provide partial incentives to participate. In that case, in the proof of Lemma 3, we can select an undominated strategy for type θ_i^0—the root of the tree T_i—that dominates the non-participation strategy $S_i^{\emptyset}$ (such a strategy exists by the definition of partial incentives to participate). Thus, we can obtain a version of Lemma 3 that assumes that Γ provides partial incentives to participate and concludes that $\min V(\Gamma)$ is upper-bounded by the value of the IC-relaxed and IR-relaxed problem. If the superior complex mechanism is required to provide full incentives to participate, then we know that all undominated strategies must dominate the non-participation strategy $S_i^{\emptyset}$. Thus, we can obtain a version of Lemma 3 which assumes that Γ provides full incentives to participate and concludes that $\min V(\Gamma)$ is upper-bounded by the value of the IC-relaxed and IR-non-relaxed problem.

A.6 Proofs for Section 4

We begin by formulating an abstract result providing sufficient conditions under which a given simple mechanism Γ is weakly dominated by a complex mechanism. Let $\mathcal{Y} \subseteq \mathcal{X}$ and define

$$\Theta_i^{\mathcal{Y}} = \left\{\theta_i \in \Theta_i : \max_{x\in\mathcal{Y}} u_i(x,\, \theta_i) > \min_{S_{-i}} u_i(g(z(h_{\emptyset},\, \boldsymbol{S}_i(\theta_i),\, S_{-i})),\, \theta_i)\right\}.$$

That is, $\Theta_i^{\mathcal{Y}}$ is the set of types of agent i that strictly prefer some outcome in $\mathcal{Y}$ to the worst possible outcome in the simple mechanism Γ.

Proposition 3. *Suppose the solution concept is obvious dominance.*[25] *Fix a simple mechanism Γ. Suppose that for some agent $i \in \mathcal{N}$, there exists $\mathcal{Y} \subseteq \mathcal{X}$ and a simple mechanism $\Gamma_{-i}^{\mathcal{Y}}$ played by agents $-i$ with an outcome space $\mathcal{Y}$ such that*

[25] In an earlier version of the paper Li ⓡ Dworczak (2022), we showed that this result also holds for the solution concepts of weak dominance and strong obvious dominance.

1. *Each outcome $x \in \mathcal{Y}$ occurs at some terminal node in $\Gamma^{\mathcal{Y}}_{-i}$;*
2. *For any $\theta_i \in \Theta_i^{\mathcal{Y}} \neq \emptyset$, the designer prefers (strictly for some $\theta_i \in \Theta_i^{\mathcal{Y}}$) the conditional expected payoff from the mechanism $\Gamma^{\mathcal{Y}}_{-i}$ to the conditional expected payoff from the mechanism Γ.*

Then, the mechanism Γ is weakly dominated.

The proof can be found in Appendix A.6.1. The idea is straightforward: Given some initial simple mechanism Γ, agent i is offered an additional option that guarantees herself an outcome in $\mathcal{Y}$. If this option is not chosen, play proceeds as in the original mechanism Γ. The key difference between a player and the designer when evaluating the additional option is that the player is only assumed to avoid obviously dominated strategies while the designer is an expected-payoff maximizer. The agent will not rule out a strategy that gives her a high enough payoff in some scenario, no matter how improbable that scenario is. The designer can structure the set $\mathcal{Y}$ and the mechanism $\Gamma^{\mathcal{Y}}_{-i}$ in a way that gives her a high payoff on average, while guaranteeing at least one contingency with a good outcome for player i.

The above construction bears some high-level resemblance to ideas first considered by Börgers and Smith (2012) and Börgers (2017) in the context of specific design environments. Given an optimal dominant-strategy mechanism, these papers introduce additional options for players in such a way that the Bayesian Nash equilibria of the resulting new mechanism provide weakly higher payoffs to the designer regardless of players' higher-order beliefs. The intuition behind Proposition 3 is indeed similar; however, Proposition 3 applies to a different solution concept and does not rely on fixing any particular environment or objective function.

A.6.1 Proof of Proposition 3

We show that the simple mechanism Γ is weakly dominated by explicitly constructing a complex mechanism that dominates it. The mechanism we construct can be interpreted as a delegation mechanism in which one agent is delegated to choose a simple mechanism (for the agents to play) from a menu of two simple mechanisms specified by the designer.[26]

Fix a player i such that the conditions in Proposition 3 are satisfied. We add a new node for player i from which play begins. Player i chooses from two options: "Choose Γ" or "Choose $\Gamma^{\mathcal{Y}}_{-i}$." If she chooses Γ, the game tree is the one associated with Γ. If she chooses $\Gamma^{\mathcal{Y}}_{-i}$, the game tree is the one associated with the game $\Gamma^{\mathcal{Y}}_{-i}$. We call this new composite game Γ'.

[26] These complex mechanisms are "type 1 strategically simple" according to the notion in Börgers and Li (2019).

First, we claim that all players other than i have an obviously dominant strategy in Γ'. This is immediate from the fact that each player $-i$ has an obviously dominant strategy in Γ and an obviously dominant strategy in $\Gamma_{-i}^{\mathcal{Y}}$. Second, all types of player i not in $\Theta_i^{\mathcal{Y}}$ also have an obviously dominant strategy which is to choose Γ, and then follow the same strategy $\boldsymbol{S}_i(\theta_i)$ that was obviously dominant for Γ. This is immediate from the fact that, for these types, the best that can happen after choosing $\Gamma_{-i}^{\mathcal{Y}}$ cannot be better than the worst possible outcome in Γ when they follow $\boldsymbol{S}_i(\theta_i)$. Third, we claim that, for all types $\theta_i \in \Theta_i^{\mathcal{Y}}$, the option to choose $\Gamma_{-i}^{\mathcal{Y}}$ is not obviously dominated. Indeed, fix the strategy profile S_{-i} that yields the minimum in the definition of $\Theta_i^{\mathcal{Y}}$, and—using condition 1 in the proposition—let $S_{-i}^{\mathcal{Y}}$ be the profile that leads to the outcome $x^\star \in \operatorname{argmax}_{x\in\mathcal{Y}} u_i(x, \theta_i)$ in the game $\Gamma_{-i}^{\mathcal{Y}}$. Then, if players $-i$ follow the strategy $(S_{-i}, S_{-i}^{\mathcal{Y}})$ in Γ', by definition of $\Theta_i^{\mathcal{Y}}$, the best response for type θ_i is to choose the game $\Gamma_{-i}^{\mathcal{Y}}$ at the first decision node.

Overall, it is not obviously dominated for types $\theta_i \in \Theta_i^{\mathcal{Y}}$ to choose the game $\Gamma_{-i}^{\mathcal{Y}}$, and when they do, play among players $-i$ in that subgame proceeds as in the original game $\Gamma_{-i}^{\mathcal{Y}}$. By condition 2 in the proposition, the designer receives a higher (sometimes strictly) conditional expected payoff in that case, compared to the conditional expected payoff she would have received in the game Γ. Therefore, the game Γ is weakly dominated by Γ'.

A.6.2 Proof of Claim 1

We first consider the construction for the case $N = 2$ without relying on randomization. We then cover the case $N = 1$.

Fix a simple mechanism Γ, with some expected revenue R. Suppose that $N \geq 2$. For any i, define $\mathcal{Y} \subset \mathcal{X}$ to contain two outcomes: *(1)* Player i pays M to the designer while some player j receives $\epsilon > 0$ from the designer, and *(2)* Player i receives M from the designer while player j receives 0 from the designer. The mechanism $\Gamma_{-i}^{\mathcal{Y}}$ is defined as having just one information node for player j who chooses between the two options in $\mathcal{Y}$. This satisfies condition 1 of Proposition 3. Moreover, the mechanism $\Gamma_{-i}^{\mathcal{Y}}$ is simple because player j has an obviously dominant strategy to select option *(1)*—player i pays M to the designer while player j receives $\epsilon > 0$ from the designer. When M is large enough, $\Theta_i^{\mathcal{Y}} = \Theta_i$, and when additionally ϵ is small enough, condition 2 of Proposition 3 holds, since the conditional expected payoff from the mechanism $\Gamma_{-i}^{\mathcal{Y}}$ is unbounded in M. Thus, by Proposition 3, the simple mechanism Γ is weakly dominated. It remains to check that the weakly-dominating complex mechanism provides partial incentives to participate. This is immediate from the proof of Proposition 3 that explicitly constructs the weakly-dominating complex mechanism Γ': Intuitively, in Γ', player i chooses between the games Γ and $\Gamma^{\mathcal{Y}}_{-i}$; thus, since each type of player i had an obviously dominant strategy dominating non-participation in Γ, each type continues to have at least one OSP-undominated strategy dominating non-participation in Γ'.

Consider the case $N = 1$. Let $\mathcal{Y} \subset \mathcal{X}$ contain two outcomes: *(1)* Player i pays M to the designer, and *(2)* Player i receives M from the designer. The mechanism $\Gamma^{\mathcal{Y}}_{-i}$ is defined as having just one information node for Nature that chooses option *(1)* with probability $1 - \epsilon$ and option *(2)* with probability ϵ. For ϵ sufficiently small, by Proposition 3 and the same arguments as before, the mechanism Γ is weakly dominated. (Of course, this construction could also be used for general N.)

A.6.3 Proof of Claim 2

Fix Γ that is a revenue-maximizing simple mechanism. Because of symmetry, we can assume without loss of optimality for the designer that Γ is symmetric as well. Define $\mathcal{Y}$ to contain two options: *(1)* Player i wins the object and pays $\bar{u}$ while some player j receives $\epsilon > 0$, and *(2)* Player i wins the object and pays $\bar{u} - \delta$ while player j receives 0. In the game $\Gamma^{\mathcal{Y}}_{-i}$, player j selects one option from $\mathcal{Y}$. $\Gamma^{\mathcal{Y}}_{-i}$ is thus simple, with player j always selecting option *(1)*. Take $\delta > 0$ small enough so that $\Theta_i \cap (\bar{u} - \delta, \bar{u}) = \emptyset$ (using finiteness of the type space). Then, we have $\Theta^{\mathcal{Y}}_i = \{\bar{u}\}$, because for all other types the payoff from $\mathcal{Y}$ is strictly negative (while the original strategy in Γ guarantees a non-negative payoff).[27] Moreover, type $\bar{u}$ always gets a non-negative payoff in $\Gamma^{\mathcal{Y}}_{-i}$ so she has a full incentive to participate. Condition 1 of Proposition 3 holds, while condition 2 is satisfied as long as the mechanism Γ extracts less than $\bar{u}$ from type $\bar{u}$ of player i. Thus, by Proposition 3, Γ is weakly dominated as long as it is not a mechanism that offers a price $\bar{u}$ to player i. By symmetry, the only case not covered by the argument is when Γ is payoff-equivalent to a mechanism that offers a price $\bar{u}$ to all players in random order. However, such a mechanism is suboptimal by assumption.

A.6.4 Proof of Claim 3

By assumption, the unique optimal simple mechanism is outcome-equivalent to offering some price $p^\star$; let $\bar{v}$ denote the corresponding optimal expected revenue. Note that, by optimality, there must exist a type $\theta = p^\star$ in Θ.

Towards a contradiction, suppose that there exists a weakly dominating complex mechanism Γ. By definition, $\min V(\Gamma) \geq \bar{v}$; we will first show that $\min V(\Gamma) = \bar{v}$. We

[27] $\bar{u} \in \Theta^{\mathcal{Y}}_i$ as long as there is positive probability that she does not receive the object in Γ—that is guaranteed by the assumption that the players and Γ are symmetric.

claim that there exists a selection from the set of OSP-undominated strategies $U_i(\cdot)$ in Γ such that *(i)* if every type plays the assigned strategy, the expected revenue for the designer is equal to at least $\min V(\Gamma)$, and *(ii)* local downward incentive constraints hold, that is, the strategy assigned to type θ_i obviously dominates (for θ_i) the strategy assigned to the highest type lower than θ_i. This follows from Lemma 3 in Appendix A.5 that establishes a more general result. Moreover, it is well known that for revenue maximization with a single player, only local downward incentive constraints bind in the optimal mechanism; hence, the expected revenue under the selection of undominated strategies described above is upper-bounded by $\bar{v}$. Since $\bar{v} \leq \min V(\Gamma)$, we conclude that $\min V(\Gamma) = \bar{v}$ and there exists a selection from the set of undominated strategies which satisfies local downward incentive constraints and yields an expected revenue of $\bar{v}$.

By the assumption that the optimal simple mechanism is unique, the only way to generate the expected revenue of $\bar{v}$ while satisfying local downward incentive constraints is for all types weakly above $p^\star$ to buy for sure at the expected price of $p^\star$. Because of full incentive to participate, type $p^\star$ must have a strategy under which she buys for sure at a price of $p^\star$ in every history. In particular, Γ must offer such a strategy. Moreover, again because of full incentives to participate, this strategy must be obviously dominant for type $p^\star$.

By the assumption that Γ weakly dominates the simple mechanism, there must exist a strategy S_1 that is OSP-undominated for some type θ, and a strategy for Nature S_0 such that the agent pays $q > p^\star$ in the outcome $g(z(h_\emptyset, S_0, S_1))$. Because the strategy "buy for sure at a price of $p^\star$" is available, for S_1 to be OSP-undominated, it must sometimes (for some strategy of Nature) generate an outcome "buy with probability x at a price of r" that is strictly preferred by θ to the outcome "buy for sure at a price of $p^\star$." At the same time, type $p^\star$ cannot derive positive utility from the outcome "buy with probability x at a price of r" as otherwise the strategy S_1 would be undominated for type $p^\star$, violating full incentives to participate (since S_1 sometimes leads to the agent paying $q > p^\star$). Moreover, note that $\theta > p^\star$, because of full incentives to participate. We are ready to obtain a contradiction: By the above reasoning, we have $\theta x - r > \theta - p^\star$ and $p^\star x - r \leq 0$. This implies

$$\theta - p^\star < \theta x - r \leq \theta x - p^\star x = x(\theta - p^\star) \leq \theta - p^\star,$$

a contradiction.

A.6.5 Supplemental Material for Section 4

We show that the uniqueness assumption in Claim 3 is needed. We construct an example in which the optimal OSP mechanisms are not unique, and they are weakly dominated. The agent has value 1 or 2, with equal probabilities. The OSP mechanisms generate an expected revenue of 1 (which can be obtained by charging a price of 1, or a price of 2). The weakly dominating mechanism features Nature that plays H with some small probability $\epsilon > 0$, and T otherwise. The mechanism offers three strategies to the agent, where the first number in every cell is the probability of trade, and the second number is the payment to the designer.

	H	T
S_1	$(1, 3/2)$	$(1, 3/2)$
S_2	$(1, 1)$	$(1, 2)$
S_3	$(1/2, 1/2)$	$(1/2, 1/2)$

For type 2, S_1 and S_2 are OSP-undominated (playing S_3 can be ruled out under the assumption that the designer can choose between strategies that are payoff equivalent, in this case S_1 and S_3). Type 1 has an obviously dominant strategy S_3. Thus, Γ provides full incentives to participate. In the worst case, type 2 plays S_1, and the designer obtains an expected revenue of 1. In the best case, type 2 plays S_2, and the designer obtains an expected revenue of $1/2 + 1/2 \cdot (2 - \epsilon)$ which can be arbitrarily close to 3/2.