STRATEGICALLY SIMPLE MECHANISMS

TILMAN BÖRGERS

Department of Economics, University of Michigan

JIANGTAO LI

School of Economics, University of New South Wales

ABSTRACT. We define and investigate a property of mechanisms that we call "strategic simplicity," and that is meant to capture the idea that, in strategically simple mechanisms, strategic choices are easy. We define a mechanism to be strategically simple if strategic choices can be based on first-order beliefs about the other agents' preferences alone, and there is no need for agents to form higher-order beliefs, because such beliefs are irrelevant to agents' optimal choices. All dominant strategy mechanisms are strategically simple. But many more mechanisms are strategically simple. In particular, strategically simple mechanisms may be more flexible than dominant strategy mechanisms in the voting problem and the bilateral trade problem.

Date: November 26, 2017.

We are grateful to Gabriel Carroll, Yi-Chun Chen, Johannes Hörner, Heng Liu, Alessandro Pavan, and Satoru Takahashi for very helpful comments.

Contents

1. Introduction

In mechanism design it often seems desirable for the designer to offer a mechanism in which agents face a straightforward choice problem, and need not engage in complex thinking to determine their optimal choices. It seems more likely that agents make the choices that the designer expects them to make if strategic thinking in the mechanism is simple than when it is complicated. Also, agents may be more willing to participate in simple mechanisms. Finally, it may be desirable that the outcomes of a mechanism don't depend too much on the cognitive abilities of agents. All these arguments provide potential reasons for constructing mechanisms in which strategic choices are easy to make.

One class of mechanisms in which one might argue that it is easy to choose a strategy are dominant strategy mechanisms. In such mechanisms, agents need not think at all about the motives of the other agents, or the other agents' rationality. This is because agents have at least one strategy that is optimal regardless of what the other agents do, and they can just choose such a strategy.[1] Of course, strategic choices may be complicated for other reasons. For example, agents may fail to recognize that they have a dominant strategy. In this paper, we abstract from such problems, and focus exclusively on the difficulties that agents may face analyzing the motives and choices of other agents.[2]

For many mechanism design problems, the class of dominant strategy mechanisms is quite small, and only includes mechanisms that are rather unattractive for a mechanism designer who wants to maximize, say, revenue, or welfare.[3] The contribution of this paper is to introduce a new property of mechanisms, called "strategic simplicity," that captures the idea that agents can determine their optimal strategies without having to think too hard about the motives of the other agents. The set of strategically simple mechanisms includes, and is strictly larger than, the set of dominant strategy

[1]Here, we use the phrase "dominant strategy" in the sense in which it is used in mechanism design theory, that is, a strategy that is optimal regardless of what the other agents do. This is slightly different from a strategy that is "weakly dominant" or a strategy that is "strictly dominant" as these terms are defined in game theory.

[2]Li [22] proposes the notion of "obviously strategy-proof mechanisms," which are mechanisms in which the task of identifying dominant strategies is, in some sense, obvious. We shall compare our approach with Li's approach in Section 8.

[3]See the examples in Chapter 4 of Börgers [8].

mechanisms. Our results suggest that in applications strategically simple mechanisms may be more attractive to a mechanism designer concerned with fairness, efficiency, or revenue, than dominant strategy mechanisms.

To illustrate our idea, it is best to consider an example. Suppose that a mechanism designer wants to determine the terms of trade between two agents, a seller and a buyer. It is known from Hagerty and Rogerson [19] that the only dominant strategy mechanisms that are ex post budget balanced and individually rational are posted price mechanisms. In a posted price mechanism, the designer chooses a (possibly random) price, without taking into account any of the agents' private information, and then agents decide whether to agree or not to agree to trade at this price. Trade comes about only when both agents agree. Obviously, this is a rather unappealing mechanism for a welfare maximizing mechanism designer.

Now consider an alternative mechanism that we call "price cap mechanism." The mechanism designer sets a price cap, but allows the seller to reduce the price. The buyer then decides whether or not to trade at this potentially reduced price. The seller clearly does not have a dominant strategy. Whether or not to reduce the price, and how far to reduce the price, depends on the seller's belief about the buyer's willingness to pay. But, regardless of her belief, the seller will never reduce the price below her reservation value, and the buyer will never agree to trade if the potentially reduced price is above his willingness to pay. In comparison to the posted price mechanism, this mechanism facilitates more efficient trade.[4]

In the price cap mechanism, the buyer faces a straightforward choice problem. The buyer agrees to trade if and only if his willingness to pay is weakly higher than the price offered. The seller's problem is arguably not too complicated either. If she believes that the buyer accepts the trade if and only if his willingness to pay is weakly higher than the price offered, then all that she needs to do is to consider her belief about the buyer's willingness to pay. This problem is equivalent to the standard monopoly problem with a price ceiling as taught in undergraduate microeconomics. For any belief that the seller might have, it is a straightforward optimization problem. Our formal definition of strategic simplicity will imply that the price cap mechanism is strategically simple.

[4]This mechanism was discussed in Börgers and Smith [9].

On the other hand, the double auction as described in Chatterjee and Samuelson [10] is, in our terminology, not strategically simple. To see why, note that in the double auction, the seller has to form her belief about the price that the buyer offers. Ideally, she would like to ask for a price that is as close as possible to, but below the price that is offered by the buyer, provided that this price is above her reservation value. But to form her belief about the price that the buyer offers, presumably the seller first has to form her belief about the buyer's belief about the seller's reservation value. Similarly, the buyer has to form his belief about the seller's belief about the buyer's willingness to pay. Potentially, infinitely many layers of such beliefs matter. Mechanisms that require of agents this level of depth of thinking will, in our terminology, not be strategically simple.

Motivated by these examples, we shall define in this paper a mechanism to be strategically simple if optimal choices can be determined using first-order beliefs alone, and there is no need for agents to form higher-order beliefs because such beliefs are irrelevant to agents' optimal choices. Here, we are referring to beliefs about the other agents' utility functions and rationality. Thus, a "first-order belief" of agent i is agent i's belief about the other agents' ($j \neq i$) utility functions, and about the other agents' rationality. "Higher-order beliefs" are, for example, agent i's belief about agent j's belief about agent i's utility function, and about agent j's belief about agent i's rationality. We shall call a mechanism strategically simple if for each agent i, her belief about the other agents' ($j \neq i$) utility functions, combined with certainty that the other agents are rational, imply which choices are optimal for agent i.

Forming beliefs about other agents' utility functions and rationality, beliefs about beliefs, beliefs about beliefs about beliefs, etc., seems to be the core of "strategic thinking." That belief formation is a costly process has been argued, for example, by Binmore [6, pp. 129-132], who argues that achieving the consistency that is necessary for having a well-defined prior requires many costly iterations of attempted belief formation.[5] Kneeland [20] offers experimental evidence that roughly 30% percent of subjects in experiments don't even form second or higher-order beliefs.[6] One may speculate that Kneeland's subjects find the cost of forming higher-order beliefs

[5]Binmore asserts that in many circumstances, including games, it is actually impossible to carry out this process.

[6]Lim and Xiong [23] have a similar finding.

prohibitive. Strategically simple mechanisms allow agents to economize on belief formation costs. A planner concerned with welfare will find this attractive. Even a planner who does not care about agents' belief formation costs per se will find strategically simple mechanisms attractive because agents' higher-order beliefs might be hard to predict.

Strategic simplicity can also be interpreted as a form of robustness in the sense of Bergemann and Morris [4]. Whereas Bergemann and Morris study implementation that does not rely on any conditions on agents' hierarchies of beliefs, we study implementation of outcomes that may depend on agents' first layer of beliefs, but not on any higher-order beliefs.

Our definition of strategic simplicity allows for the possibility that only some subset of all utility functions and only some subset of all first-order beliefs is considered. Our main result shows that, under a "richness" condition on the domain of utility functions and beliefs, strategic simplicity is equivalent to a "local dictatorship" property of the mechanism. In contrast with (classical) dictatorship, local dictatorship means, roughly speaking, that there is some agent who dictates the outcome if we restrict attention for every agent to certain subsets of her strategy set. The identity of the dictator may depend on the subsets that we consider. Every dictatorship is a local dictatorship, but there are many more local dictatorships than dictatorships.

Our characterization result suggests a natural division of strategically simple mechanisms into two categories: mechanisms in which there is some agent who is a local dictator at all restrictions that we consider, and mechanisms in which this is not the case. We shall call the former "type 1 strategically simple mechanisms," and the latter "type 2 strategically simple mechanisms." Type 1 mechanisms are easy to characterize. One can think of type 1 mechanisms as "delegation mechanisms:" the mechanism designer delegates the choice of the mechanism to a "delegate," who chooses a mechanism from a given set of dominant strategy mechanisms that the designer has specified. The delegate's choice will depend on her first-order belief, while the other agents' choices don't require any belief formation. Type 2 mechanisms are harder to characterize. This paper will offer examples of such mechanisms when they exist, but we do not have a complete characterization of type 2 mechanisms.

The paper is organized as follows. Section 2 contains the formal definition of strategic simplicity. Section 3 illustrates the definition with examples. In Section 4, we discuss the connection between our definition of strategic simplicity and epistemic game theory. We include a discussion of epistemic foundations because the language that we use in this Introduction to motivate our concept of "strategic simplicity" involves expressions such as "first-order belief" and "higher-order beliefs," and these are expressions taken from epistemic game theory. However, for simplicity, the definition in Section 2 does not involve any explicit reference to epistemic game theory. In Section 4, we sketch how an epistemic approach to the definition would proceed more explicitly. We also discuss the connection with rationalizability in that section. A natural conjecture is that a mechanism is strategically simple according to our definition if it can be solved in two steps of elimination of strategies that are not best responses, that is, in two steps of the procedure that defines rationalizability. In Section 4, we explain that, while there is some merit to this conjecture, the conjecture is not completely correct.[7]

Section 5 contains our characterization result of strategically simple mechanisms under a richness condition on the domain of utility functions and beliefs. Sections 6 and 7 consider the applications of our main results to the voting problem and the bilateral trade problem. The results in Section 6 demonstrate that in the voting environment, the class of strategically simple mechanisms is much larger than the class of dominant strategy mechanisms. By the celebrated Gibbard-Satterthwaite [18, 26] Theorem, in the voting environment as we define it here, a mechanism has dominant strategies and if and only if it is dictatorial. There are many more strategically simple voting mechanisms that reflect unanimous preferences. The results in Section 7 show that in the bilateral trade environment, the only strategically simple mechanisms are those in which one agent proposes terms of trade, and the other agent accepts or rejects. We discuss related literature in Section 8. Section 9 concludes.

[7]In the context of this discussion, we shall be more precise about the relevant notion of rationalizability. This will matter as the literature has developed several different notions of rationalizability in games of incomplete information.

2. Definitions

There are n agents: $i \in I = \{1, 2, \ldots, n\}$, and a finite set A of outcomes. A mechanism consists of finite strategy sets S_i for each agent i, and a function $g : \times_{i \in I} S_i \to A$ that describes for each choice of strategies which outcome will result. We define $S \equiv \times_{i \in I} S_i$ with generic element s, and, for every $i \in I$, we define $S_{-i} \equiv \times_{j \neq i} S_j$ with generic element s_{-i}. We assume that there are no duplicate strategies: for every $i \in I$, for all $s_i, s_i' \in S_i$ with $s_i \neq s_i'$, there is some $s_{-i} \in S_{-i}$ such that $g(s_i, s_{-i}) \neq g(s_i', s_{-i})$.

A von Neumann-Morgenstern (vNM) utility function of agent i is a function $u_i : A \to \mathbb{R}$. We define $\mathcal{U}$ to be the set of all utility functions such that that: $u(a) \neq u(a')$ whenever $a \neq a'$, $\min_{a \in A} u_i(a) = 0$, and $\max_{a \in A} u_i(a) = 1$. Thus we rule out indifferences and normalize utility. This simplifies arguments below. We write $u \equiv (u_1, u_2, \ldots, u_n)$ and $u_{-i} \equiv (u_j)_{j \neq i}$.

For every agent i there is a non-empty, Borel-measurable set $\mathbf{U_i} \subseteq \mathcal{U}$ of utility functions that are possible utility functions of agent i. We allow for the possibility that $\mathbf{U}_i \neq \mathcal{U}$ to be able to capture assumptions such as the assumption that agents' utility functions are quasi-linear. We define $\mathbf{U} \equiv \times_{i \in I} \mathbf{U}_i$, and, for every $i \in I$, we define $\mathbf{U}_{-i} \equiv \times_{j \neq i} \mathbf{U}_j$.

For a given mechanism, for every i and every $u_i \in \mathbf{U}_i$, we denote by $UD_i(u_i)$ the set of all strategies that are not weakly dominated for agent i with utility function u_i, where weak dominance may be by a pure or by a mixed strategy. If $u \in \mathbf{U}$, we define $UD(u) \equiv \times_{i \in I} UD_i(u_i)$, and, for every $i \in I$ and every $u_{-i} \in \mathbf{U}_{-i}$, we define $UD_{-i}(u_{-i}) \equiv \times_{j \neq i} UD_j(u_j)$. To avoid tedious detail, we assume that for every agent $i \in I$ and every strategy $s_i \in S_i$, there is at least some $u_i \in \mathbf{U}_i$ such that $s_i \in UD_i(u_i)$.

A "utility belief" μ_i of agent i is a Borel probability measure on $\mathbf{U}_{-i}$. We interpret μ_i as agent i's "first-order" belief. Higher-order beliefs would be beliefs about other agents' beliefs about utility functions, etc. As indicated in the Introduction, we want to focus on mechanisms in which higher-order beliefs play no role. Therefore, we don't formally define them here.

For any finite set (or Borel subset of a finite dimensional Euclidean space) X, we shall denote by $\Delta(X)$ the set of all (Borel) probability measures on X. The set of all possible utility beliefs of agent i is some non-empty subset $\mathbf{M}_i$ of $\Delta(\mathbf{U}_{-i})$. We allow for the possibility that $\mathbf{M}_i \neq \Delta(\mathbf{U}_{-i})$ to be able to capture assumptions such as the assumption that every agent believes

that the agents' utility functions are stochastically independent. We define $\mathbf{M} \equiv \times_{i \in I} \mathbf{M}_i$, and, for every $i \in I$, we define $\mathbf{M}_{-i} \equiv \times_{j \neq i} \mathbf{M}_j$, and we denote typical elements of these sets by μ and μ_{-i} respectively.

A "strategic belief" $\hat{\mu}_i$ of agent i is a probability measure on S_{-i}: $\hat{\mu}_i \in \Delta(S_{-i})$. Strategic beliefs are needed for agents to determine expected utility maximizing strategies. The next definition will describe how agents may derive a strategic belief from a utility belief. We assume that agents are certain that other agents do not play weakly dominated strategies. Then, loosely speaking, a strategic belief can be obtained from a given utility belief by dividing the probability assigned to any utility function u_j $(j \neq i)$ in some arbitrary way among the not weakly dominated strategies of agent j with utility function u_j. We call a strategic belief that can be derived in this way from a utility belief "compatible with the utility belief." Obviously, for given utility belief, there may be many compatible strategic beliefs. We formally define the compatibility of strategic beliefs with utility beliefs as follows:

Definition 1. *A strategic belief $\hat{\mu}_i$ is "compatible with a utility belief μ_i" if there is a probability measure ν_i on $S_{-i} \times \mathbf{U}_{-i}$ that has support in*

$$\bigtimes_{j \neq i} \{(s_j, u_j) \in S_j \times \mathbf{U}_j | s_j \in UD_j(u_j)\}$$

and that has marginals $\hat{\mu}_i$ on S_{-i} and μ_i on $\mathbf{U}_{-i}$.

In this definition, ν_i is agent i's joint belief about strategies and utility functions of the other agents. Agent i's certainty that the other agents don't play weakly dominated strategies is captured by the support restriction in Definition 1. The belief ν_i must also reflect the given utility belief μ_i of agent i, that is, ν_i's marginal on $\mathbf{U}_{-i}$ must be μ_i. The marginal on S_{-i} is then a compatible strategic belief. We denote the set of all strategic beliefs that are compatible with a given utility belief μ_i by $\mathcal{M}_i(\mu_i)$.

Given a utility function $u_i \in \mathbf{U}_i$ and a strategic belief $\hat{\mu}_i \in \Delta(S_{-i})$ of agent i, we denote by $BR_i(u_i, \hat{\mu}_i)$ the set of all strategies in $UD_i(u_i)$ that maximize expected utility in S_i.

We are now ready to provide the key definition of this paper.

Definition 2. *A mechanism is "strategically simple" if for every agent i, every utility function $u_i \in \mathbf{U}_i$, and every utility belief $\mu_i \in \mathbf{M}_i$:*

$$\bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i)} BR_i(u_i, \hat{\mu}_i) \neq \varnothing.$$

What we require here for every agent i, every utility function u_i of agent i, and every utility belief μ_i of agent i, is that agent i has at least one strategy that maximizes expected utility regardless of which strategic belief $\hat{\mu}_i$ that is compatible with the utility belief μ_i agent i picks. Thus, there is no need for agent i to try to distinguish more plausible from less plausible compatible strategic beliefs. If that was necessary, it may be helpful for agent i to form higher-order beliefs. But if a mechanism is strategically simple, there is no benefit to agent i from forming higher-order beliefs.

Often a mechanism designer's interest is not in the mechanism itself, but in the outcomes that result when agents pick their strategies rationally. For strategically simple mechanisms, which strategy maximizes expected utility will depend not only on an agent's utility function, but also on this agent's utility belief. It is therefore natural to focus on correspondences that map utility functions and utility beliefs into sets of outcomes. We call such correspondences "outcome correspondences."

Definition 3. *The "outcome correspondence" implemented by a strategically simple mechanism is the correspondence:*

$$F : \mathbf{U} \times \mathbf{M} \twoheadrightarrow A$$

defined by:

$$F(u,\mu) \equiv g\left(\bigtimes_{i\in I}\left(\bigcap_{\hat{\mu}_i\in\mathcal{M}_i(\mu_i)} BR_i(u_i,\hat{\mu}_i)\right)\right) \text{ for all } (u,\mu)\in\mathbf{U}\times\mathbf{M}.$$

The following definition will be useful:

Definition 4. *Two strategically simple mechanisms are "equivalent" if they implement the same outcome correspondence.*

The literature often refers to "social choice correspondences," which are similar to "outcome correspondences," except that their domain consists of profiles of utility functions, or preferences, only, and does not include profiles of first-order beliefs. Focusing on utility functions in the domain seems natural if one gives the correspondence a normative interpretation, as a description of the outcomes that the mechanism designer regards as desirable. Here, however, we give our correspondence a positive interpretation: it is a description of the end result of a given mechanism. By including the first-order beliefs in this description, we give a more detailed description of the consequences resulting from rational choice in a given mechanism than we

would obtain if only preferences were in the domain of the correspondences that we are considering.

An implicit assumption in our definition of outcome correspondences is that for any given utility function and utility belief agents will only choose strategies from the set $\bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i)} BR_i(u_i, \hat{\mu}_i)$. This implies that an agent i will not choose a strategy if it is a best response to only some strategic beliefs compatible with the agent's given utility belief, but not to all such strategic beliefs. This assumption is in the spirit of our basic hypothesis according to which agents find it costly to refine their strategic beliefs, beyond making it compatible with their utility belief, and will avoid doing so if they can.

One can interpret singleton-valued outcome correspondences as direct mechanisms in which agents report their utility functions and their utility beliefs. Using this interpretation, one can then ask whether a revelation principle holds, i.e.: If a singleton-valued outcome correspondence is implemented by a strategically simple mechanism, is then the direct mechanism defined by the outcome correspondence itself a strategically simple mechanism, and is truth telling an optimal strategy for all utility functions and first-order beliefs, regardless of higher-order beliefs, in this mechanism? Unfortunately, a technical problem that we encounter when asking this question is that we have defined strategically simple mechanisms only for the case that a mechanism has a finite strategy set for each agent, whereas we have allowed the sets of pairs of utility functions and beliefs to be infinite, and thus the direct mechanism may have infinite strategy sets. This problem is bypassed if attention is restricted to the case of finite $\mathbf{U} \times \mathbf{M}$. In this case, one can then verify that the revelation principle as described above holds. Some of our analysis below will, however, specifically be about the case of infinite $\mathbf{U} \times \mathbf{M}$ and finite mechanisms, and therefore the revelation principle will not play an important role in our analysis, in contrast to the conventional theory of mechanism design.

The formal framework developed in this section suggests two possible focuses for our analysis: the characterization of strategically simple mechanisms, or the characterization of the outcome correspondences that can be implemented by strategically simple mechanisms. We find it convenient to focus on mechanisms themselves. But we shall explain some of the implications of our results for implementable outcome correspondences.

3. Examples

To illustrate our notion of strategically simple mechanisms, we start by discussing three examples. In each case, agent 1 (he) and agent 2 (she) collectively choose an outcome from three alternatives $\{a, b, c\}$. We make no restrictions regarding the agents' utilities or beliefs. As we shall see, the mechanisms in Example 1 and Example 3 are strategically simple. In other words, agents can figure out the expected utility maximizing strategy on the basis of their first-order beliefs alone. The mechanism in Example 2 is not strategically simple.

Example 1. The mechanism shown in Figure 1 is strategically simple. In this mechanism, agent 1 either chooses alternative b or offers the menu $\{a, c\}$ to agent 2, who then chooses an alternative from the menu. Agent 2 has a dominant strategy: she chooses a if and only if she ranks a above c. Because agent 2 has a dominant strategy, for any utility belief of agent 1, there is a unique strategic belief that is compatible with the given utility belief. Thus, agent 1 can figure out his expected utility maximizing strategy on the basis of his first-order belief about agent 2's preference alone.

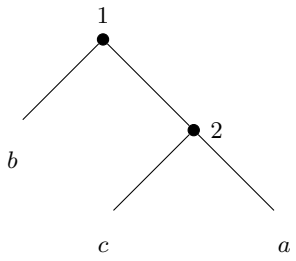

Figure 1. Example 1

Example 2. The mechanism shown in Figure 2 is not strategically simple. It suffices to consider a special case: agent 1 has preference cba,[8] and agent 2 has preference cab and attaches probability 1 to the event that agent 1 has preference cba. In this case, agent 1 has a weakly dominant choice to continue at the first decision node. At the third decision node, he needs to choose between his middle alternative b and leaving agent 2 the choice

[8]We use cba to denote the preference that the agent ranks c above b, and ranks b above a.

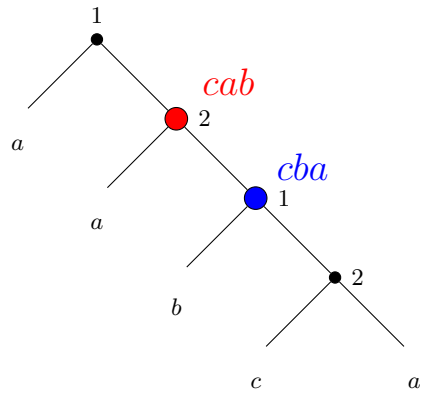

FIGURE 2. Example 2

between his best alternative c and his worst alternative a. Note that both actions are not weakly dominated. Agent 2 with preference cab at the second decision node is choosing between her middle alternative a and leaving it to agent 1 to choose agent 2's worst alternative b or to give agent 2 the chance to pick her best alternative c. The probability of getting her best alternative depends on the choice of agent 1 at the third decision node, and agent 2 does not have a strategy that is a best response to all strategic beliefs that are compatible with her utility belief. Therefore, agent 2 cannot determine her choice at the second decision node on the basis of first-order belief alone. Or can also show that agent 1 cannot determine his choice at the third decision node on the basis of first-order belief alone. The mechanism is clearly not strategically simple for the agents.

Example 3. The mechanism shown in Figure 3 is strategically simple. In what follows, we show that agents can determine their expected utility maximizing strategies on the basis of their first-order beliefs alone.

This is obvious if an agent has a weakly dominant strategy. If agent 1 ranks a highest, he has a weakly dominant choice to stop at the first decision node. If agent 1 ranks b highest, he has a weakly dominant choice to continue at the first decision node and stop at the third decision node. If agent 1 has preference cab, he has a weakly dominant choice to continue at the first, third, and fifth decision nodes.

If agent 2 ranks a highest, she has a weakly dominant choice to stop at the second decision node. If agent 2 ranks b highest, she has a weakly dominant choice to continue at the second decision and the fourth decision

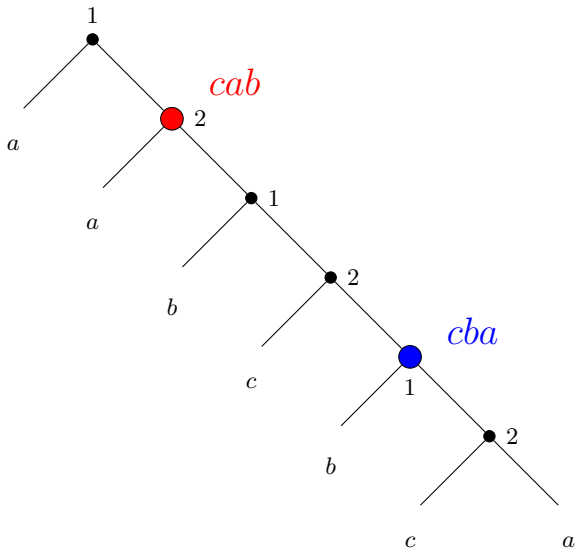

FIGURE 3. Example 3

node, and she chooses a at the sixth decision node if her preference is bac and c if her preference is bca. If agent 2 has preference cba, she has a weakly dominant choice to continue at the second decision node and stop at the fourth decision node.

We have thus only two cases in which there are multiple not weakly dominated strategies. In the first case, agent 1's preference is cba. Agent 1 with preference cba has a weakly dominant choice to continue at the first decision node and the third decision node. But at the fifth decision node, he needs to choose between his middle alternative b and a lottery on his best alternative c and worst alternative a. The weight of the lottery depends only on agent 1's first-order belief about agent 2's preferences. To see this, note that if agent 2 ranks a or c highest, the game would have stopped by the fourth decision node. If agent 2's preference is bac or bca, she would choose to continue at the second decision node and the fourth decision node. Without loss of generality, we assume that agent 1's utility function is: $u_1(c) = 1, u_1(b) = x, u_1(a) = 0$, and that he attaches probability p (resp. q) to the event that agent 2's preference is bac (resp. bca). Agent 1 chooses to stop at the fifth decision node if and only if $x \geqslant \frac{q}{p+q}$. In other words, agent 1 can determine his expected utility maximizing strategy at the fifth decision node on the basis of his first-order belief about agent 2's preferences alone.

The second case in which there are multiple undominated strategies is that agent 2 has preference *cab*. Without loss of generality, we assume that agent 2's utility function is: $u_2(c) = 1, u_2(a) = y, u_2(b) = 0$, and that she attaches probability $\tilde{p}$ to the event that agent 1 ranks b highest, in which case agent 1 stops at the third decision node, and probability $\tilde{q}$ to the event that agent 1 ranks c highest, in which case agent 1 continues at the third decision node. Then agent 2 will choose to stop at the second decision node if and only if: $y \geqslant \frac{\tilde{q}}{\tilde{p}+\tilde{q}}$. Agent 2 has a weakly dominant choice to stop at the fourth decision node. Again, we see that agent 2's optimal choice only depends on her first-order belief, as required by strategical simplicity. This completes the argument that the mechanism in Figure 3 is strategically simple.

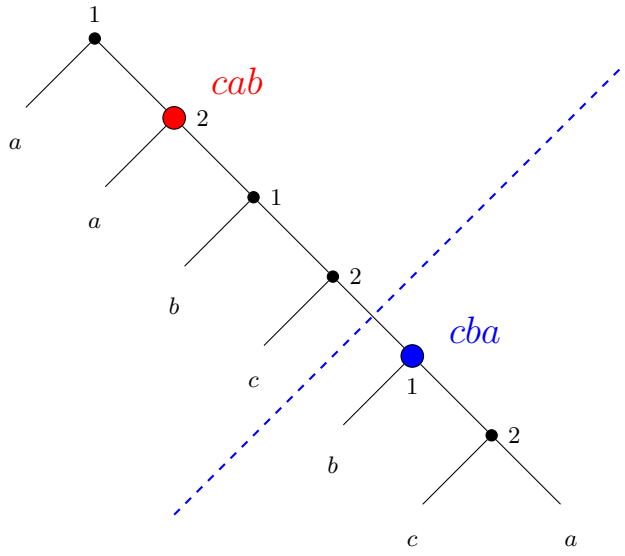

FIGURE 4. Example 3 revisited

As discussed above, agent 1's first-order belief matters at the fifth decision node if his preference is *cba*, and agent 2's first-oder belief matters at the second decision node if her preference is *cab*. One might expect that agent 2 needs to formulate her belief about agent 1's first-order belief to determine the expected utility maximizing strategy, and vice versa. We note that there is the following separation feature in the mechanism, indicated by the dashed line in Figure 4. Although agent 1's belief matters at the fifth decision node, agent 2 with preference *cab* does not have to form a higher-order belief because in any case, the game would have stopped after the

fourth node, and she does not need to take into account agent 1's choice at the fifth decision node. In other words, agent 1's choice at the fifth decision node does not have any influence on agent 2's choice at the second decision node if agent 2 has preference *cab*.

4. Strategic Simplicity and Epistemic Game Theory

Our intuitive discussion of strategic simplicity in the Introduction invokes the concepts of first and higher-order beliefs about other agents' utility functions and about their rationality. This suggests that our definition of strategic simplicity should be describable in the language of epistemic game theory, and also that there should be a connection between rationalizability and our definition of strategic simplicity. In this section, we discuss these two issues.

We first consider epistemic foundations of our approach. One might conjecture that our definition of strategic simplicity is equivalent to the requirement that for every agent i, every utility function $u_i \in \mathbf{U}_i$ and every utility belief $\mu_i \in \mathbf{M}_i$ there is just one strategy choice of agent i that is compatible with the following hypotheses: (i) agent i is an expected utility maximizer; (ii) agent i has utility function u_i; (iii) agent i's beliefs about other agents' utility functions is given by μ_i; and (iv) agent i believes with certainty that all other agents are expected utility maximizers.

If strategic simplicity were defined using the requirement just described, then it would be straightforward to show that a mechanism is strategically simple if and only if for every agent i with utility function u_i and utility belief μ_i the set:

$$\arg\max_{s_i \in S_i} \sum_{s_{-i} \in S_{-i}} u_i(g(s_i, s_{-i}))\hat{\mu}_i(s_{-i})$$

has the same, single element for all strategic beliefs $\hat{\mu}_i$ for which there is a probability measure ν_i on $S_{-i} \times \mathbf{U}_{-i}$ that has marginal μ_i on $\mathbf{U}_{-i}$, marginal $\hat{\mu}_i$ on S_{-i}, and that has support in:

$$\bigtimes_{j \neq i} \{(u_j, s_j) \in \mathbf{U}_j \times S_j | s_j \in RAT_j(u_j)\}.$$

Here, we define:

$$RAT_j(u_j) \equiv \bigcup_{\hat{\mu}_j \in \Delta(S_{-j})} \arg\max_{s_j \in S_j} \sum_{s_{-j} \in S_{-j}} u_j(g(s_j, s_{-j}))\hat{\mu}_j(s_{-j}).$$

But notice that this alternative definition of strategic simplicity differs from ours in two ways. The first is that, in our definition, "rationality" means not only expected utility maximization, but also as not playing a weakly dominated strategy.[9] As a consequence, according to our definition more mechanisms are strategically simple than would be the case if we used the alternative definition. The second way in which our definition differs from the alternative one above is that, rather than requiring the best response sets to always have the same, single element for all strategic beliefs compatible with a given utility belief, we have required that the sets have at least one element in common. This, too, implies that more mechanisms are strategically simple according to our definition than would be the case if we used the alternative definition.

Our reason for ruling out weakly dominated strategies is pragmatic. Some of our examples are most naturally interpreted as the normal forms of extensive form games, and by ruling out weakly dominated strategies we rule out some strategies that violate the most basic versions of sequential rationality. Without ruling such strategies out, the examples that we discuss would not be strategically simple, but with our definition, they are strategically simple.[10] The second difference between our definition of strategic simplicity in this paper, and the alternative definition outlined above, allows us to call mechanisms strategically simple even if there is, say, one strategy that is optimal for all strategic beliefs that are compatible with a given utility belief, but sometimes, perhaps for some knife-edge beliefs, some other strategy is optimal as well. It seems in the spirit of our approach to assume that agents who find that some strategy is optimal for all relevant beliefs will not care if there are other strategies that are optimal for only some compatible beliefs but not for others.

We now turn to the relation between our definition of strategic simplicity and rationalizability. This connection is more complicated to describe for the definition of strategic simplicity that we use in this paper than for the

[9]That is, whenever in the alternative definition the "arg max" operator appears, in our definition attention is restricted to not weakly dominated strategies.

[10]The primary objection against the use of weakly dominated strategies, that the order of elimination of weakly dominated strategies matters, does not arise in our setting, because we do not iterate the elimination of weakly dominated strategies. We could try to provide an epistemic interpretation of our use of weak dominance, perhaps along the lines of Frick and Romm [17], but we have not pursued this.

alternative definition described above. Therefore, we first consider the alternative definition. Had we used this alternative condition, then strategic simplicity would be equivalent to the requirement that the elimination procedure that defines rationalizability stops after two steps because for every type there is only one rationalizable strategy left over, that is, that the game is "rationalizability solvable" in two steps. Here, we think of rationalizability as a solution concept that conditions for each agent i on agent i's utility function u_i, and agent i's first-order belief about other agents' utility function, μ_i. Researchers have proposed several notions of rationalizability for incomplete information games. The concept of rationalizability that we just described informally can be formally defined as a special case of interim correlated rationalizability (Dekel et. al. [16, p. 20]) or of Δ-rationalizability (Battigalli [3]). We omit the details.

For a statement that relates rationalizability to the concept of strategic simplicity that we actually use in this paper, we need to modify rationalizability in two ways. In the first round, weakly dominated strategies would have to be removed, not just strictly dominated strategies. Also, for the second round we would have to require not that only a single strategy is left over, but that agents are indifferent between all surviving strategies, if they evaluate these strategies using their first-order beliefs and the conclusions of the first round.

5. Characterization

We now provide a useful characterization of strategically simple mechanisms using a richness assumption regarding the sets of relevant utility functions and beliefs. We denote by $\mathcal{R}$ the set of all linear (that is: complete, transitive, and anti-symmetric) orders on the set of alternatives A. A generic element of $\mathcal{R}$ will be denoted by R_i where the lower index refers to an agent i. Every utility function $u_i \in \mathcal{U}$ induces a linear order R_i in the following way: $aR_ib \Leftrightarrow u_i(a) > u_i(b)$. Let us denote by $\mathcal{U}(R_i)$ the set of all utility functions in $\mathcal{U}$ that induce R_i.

Next, we extend the notion of weak dominance to the case that only pure strategy dominance is considered. In this case, only the order R_i induced by agent i's utility function u_i matters.

Definition 5. *Let $R_i \in \mathcal{R}$. A strategy $s_i \in S_i$ is called "weakly dominated given R_i" if there is another strategy $\hat{s}_i \in S_i$ such that for all $s_{-i} \in S_{-i}$*

$$g(\hat{s}_i, s_{-i}) R_i g(s_i, s_{-i}),$$

and, for some $s_{-i} \in S_{-i}$

$$g(\hat{s}_i, s_{-i}) R_i g(s_i, s_{-i}) \text{ and } g(\hat{s}_i, s_{-i}) \neq g(s_i, s_{-i}).$$

We denote by $UD_i(R_i) \subseteq S_i$ the set of all strategies of agent i that are not weakly dominated given R_i.

For any list of linear orders $R = (R_1, R_2, \ldots, R_n) \in \mathcal{R}^n$ we define for every $i \in I$: $UD_{-i}(R_{-i}) \equiv \times_{j \neq i} UD_j(R_j)$.

Theorem 1. *Suppose for every agent i there is a non-empty set $\mathcal{R}_i \subseteq \mathcal{R}$ such that $\mathbf{U}_i = \bigcup_{R_i \in \mathcal{R}_i} \mathcal{U}(R_i)$, and suppose $\mathbf{M}_i = \Delta(\mathbf{U}_{-i})$ for all $i \in I$. Then a mechanism is strategically simple if and only if for every $R \in \times_{i \in I} \mathcal{R}_i$ there is an agent $i^* \in I$ such that for every strategy $s_{i^*} \in UD_{i^*}(R_{i^*})$ there is an alternative $a \in A$ such that:*

$$g(s_{i^*}, s_{-i^*}) = a \text{ for all } s_{-i^*} \in UD_{-i^*}(R_{-i^*}).$$

In words, the condition that is necessary and sufficient for strategic simplicity says the following. Whenever we fix a vector of preferences $(R_1, R_2, \ldots, R_n) \in \times_{i \in I} \mathcal{R}_i$ and consider the mechanism restricted to the strategy sets $UD_i(R_i)$ for all $i \in I$, then, in the restricted mechanism, some agent i^* is a dictator. That is, for each of the alternatives that are possible when agents choose their strategies from $UD_i(R_i)$, agent i^* has an action that enforces that alternative if all other agents choose from $UD_i(R_i)$, and each of agent i^*'s actions enforces some alternative. We call agent i^* a "local dictator," because in the restricted game agent i^* dictates which alternative is chosen.

The theorem applies only to certain domains of utility functions and beliefs. Specifically, the theorem assumes that for each agent the set of relevant utility functions is the set of all utility functions that induce some linear order from a given set of linear orders, and that for each agent the relevant beliefs are all beliefs that have support in the set of considered utility functions. We thus allow restricted domains of strategic simplicity, but domains that still satisfy strong "richness" conditions. In some settings, such as voting settings, these assumptions may be plausible, whereas in other settings, they may be less desirable. For example, when an allocation of money is part of the specification of alternatives, our assumption on the set of utility

functions considered rules out that only risk neutral agents are considered, even though that is a popular case in the mechanism design literature. The assumption on the set of relevant beliefs rules out that each agent regards the others' agents preferences as stochastically independent. Our proof of Theorem 1 makes strong use of these assumptions, and we have not yet found useful results for smaller domains.

We now describe an implication of Theorem 1 for outcome correspondences that can be implemented by strategically simple mechanisms. Informally speaking, the implemented set of alternatives can depend on at most one agent's vNM utilities and utility beliefs when we hold a preference profile R fixed and assume that all agents believe with probability 1 that the other agents have the preferences given by R. We formalize this property in the following definition. In this definition we say that a profile of utility functions u_{-i} induces a profile of preference relations R_{-i} if for every $j \in I \backslash \{i\}$ u_j induces R_j, and that a profile of utility functions u induces a profile of preference relations R if for every $j \in I$ u_j induces R_j.

Definition 6. *Let $i \in I$ and $R \in \mathcal{R}^n$. An outcome correspondence $F : \mathbf{U} \times \mathbf{M} \twoheadrightarrow A$ is "non-responsive to the vNM utilities and utility beliefs of agents $j \neq i$ at R" if, whenever $u_i \in \mathbf{U}_i$ represents R_i, $u_{-i}, \hat{u}_{-i} \in \mathbf{U}_{-i}$ both represent R_{-i}, $\mu_i \in \mathbf{M}_i$ and $\mu_{-i}, \hat{\mu}_{-i} \in \mathbf{M}_{-i}$ then:*

$$F((u_i, u_{-i}), (\mu_i, \mu_{-i})) = F((u_i, \hat{u}_{-i}), (\mu_i, \hat{\mu}_{-i})).$$

In words, the outcome correspondence is non-responsive to agents $j \neq i$ at R if, as long as agents' utility functions represent the preferences in R, then the von Neumann Morgenstern utility functions and beliefs of agents $j \neq i$ have no impact on the set of outcomes implemented. The following result follows directly from Theorem 1. We don't give a formal proof.

Corollary 1. *Suppose for every agent i there is a non-empty set $\mathcal{R}_i \subseteq \mathcal{R}$ such that $\mathbf{U}_i = \bigcup_{R_i \in \mathcal{R}_i} \mathcal{U}(R_i)$, and suppose $\mathbf{M}_i = \Delta(\mathbf{U}_{-i})$ for all $i \in I$. If an outcome correspondence $F : \mathbf{U} \times \mathbf{M} \twoheadrightarrow A$ can be implemented by a strategically simple mechanism, then for every preference profile $R \in \mathcal{R}^n$ there is some agent i^* such that the correspondence F is non-responsive to agents $j \neq i^*$ at R.*

Agent i^* in this Corollary is obviously the local dictator at R. This corollary implies, for example, that it is impossible to find a strategically

simple mechanism that on its whole domain implements alternatives that maximize ex post utilitarian welfare, that is, the sum of agents' utilities.

To obtain a further understanding of strategically mechanisms, we now partition the set of all mechanisms that are strategically simple on domains that satisfy the assumptions of Theorem 1 into two subsets. If the assumptions of Theorem 1 hold, then, for any $R \in \mathcal{R}_1 \times \mathcal{R}_2 \times \ldots \times \mathcal{R}_n$, we denote by $I^*(R)$ the set of local dictators at R.

Definition 7. *Suppose for every agent i there is a non-empty set $\mathcal{R}_i \subseteq \mathcal{R}$ such that $\mathbf{U}_i = \bigcup_{R_i \in \mathcal{R}_i} \mathcal{U}(R_i)$, and suppose $\mathbf{M}_i = \Delta(\mathbf{U}_{-i})$ for all $i \in I$. Then a strategically simple mechanism is of type 1 if:*

$$\bigcap_{R \in \times_{i \in I} \mathcal{R}_i} I^*(R) \neq \varnothing.$$

Otherwise, it is of "type 2."

In words, in a type 1 strategically simple mechanism there is an agent who is local dictator at all preference profiles, whereas that is not the case for type 2 strategically simple mechanisms.

Type 1 strategically simple mechanisms can be easily characterized. To state this characterization, we first introduce a class of mechanisms that we refer to as "delegation mechanisms."

Definition 8. *A mechanism is a "delegation mechanism" if it is the normal form of an extensive form mechanism of the following type: First, some agent $i^* \in I$ chooses an element s_{i^*} from some finite set S_{i^*}. All agents observe s_{i^*}. Then, for every s_{i^*}, a subgame with simultaneous moves follows in which the players are the agents in $I \backslash \{i^*\}$, and in which a dominant strategy mechanism with outcomes in A is played, where the mechanism may depend on s_{i^*}.*

In a delegation mechanism the mechanism designer thus delegates the choice of the mechanism to some agent i^*. This agent has to choose a mechanism from a given set of dominant strategy mechanisms that the mechanism designer has specified. Clearly, in a delegation mechanism, all agents except i^* have dominant strategies, and therefore do not even have to form first-order beliefs, and for agent i^* therefore only first-order belief are relevant to his or her choice.

Theorem 2. *Suppose for every agent i there is a non-empty set $\mathcal{R}_i \subseteq \mathcal{R}$ such that $\mathbf{U}_i = \bigcup_{R_i \in \mathcal{R}_i} \mathcal{U}(R_i)$, and suppose $\mathbf{M}_i = \Delta(\mathbf{U}_{-i})$ for all $i \in I$.*

(i) *Every delegation mechanism is a type 1 strategically simple mechanism.*
(ii) *For every type 1 strategically simple mechanism, there is an equivalent delegation mechanism.*

In the two applications that we consider in the next two sections, it will be interesting and straightforward to characterize class 1 strategically simple mechanisms. When they exist, we shall also provide examples of type 2 strategically simple mechanisms.

6. Voting

We begin our analysis of strategically simple mechanisms with the case in which no restrictions are assumed regarding agents' utilities and beliefs: $\mathbf{U}_i = \mathcal{U}$ and $\mathbf{M}_i = \Delta(\mathcal{U}^{n-1})$ for all $i \in I$. This is the most demanding form of strategic simplicity. We call a mechanism that is strategically simple on this domain a "strategically simple voting mechanism" because the unrestricted domain is a domain that has been considered in parts of the voting literature. The celebrated Gibbard-Satterthwaite [18, 26] Theorem shows that in the voting environment, a mechanism has dominant strategies if and only if it is dictatorial. As we shall see, there are many more strategically simple voting mechanisms.

The voting environment satisfies the domain assumptions from the previous section, and Theorem 1 can be applied. We shall distinguish type 1 and type 2 strategically simple voting mechanisms. In a type 1 strategically simple voting mechanism, some agent i^* chooses a subset of the set A of alternatives and a dominant strategy mechanism for the other agents to pick one alternative from this set. In a second stage, the other agents then play this dominant strategy mechanism. The influence of the first agent on the ultimate outcome may be restricted by limiting the set of subsets of A and dominant strategy mechanisms she can choose from. Example 1 in Section 3 is a type 1 strategically simple voting mechanism.

Standard results in voting theory provide characterizations of the dominant strategy mechanisms that can be picked in the second stage. If agent i^* rules out all but two alternatives, then a mechanism has dominant strategies if and only if it is a generalized form of majority voting (Barberà [2, p. 759]). If agent i^* allows the other agents to pick from at least three alternatives, then only dictatorial mechanisms have dominant strategies, by

the Gibbard-Satterthwaite theorem. Thus, agent i^*, if she wants to allow at least three alternatives, then has to pick one of the other agents, and needs to let this agent make the ultimate decision, where this agent is restricted to the set of alternatives chosen by agent i^*.

We do not have a characterization of type 2 strategically simple voting mechanisms. The mechanism in Example 3 is a type 2 voting mechanism. In what follows, we shall present another example of type 2 voting mechanism. These two mechanisms are the only type 2 strategically simple voting mechanisms that we have identified so far. While these examples are intriguing, we have not yet established a general pattern that would allow us to find a large class of examples, let alone a class that exhausts the set of all type 2 strategically simple mechanisms. It would be desirable to obtain a characterization of type 2 strategically simple voting mechanisms.

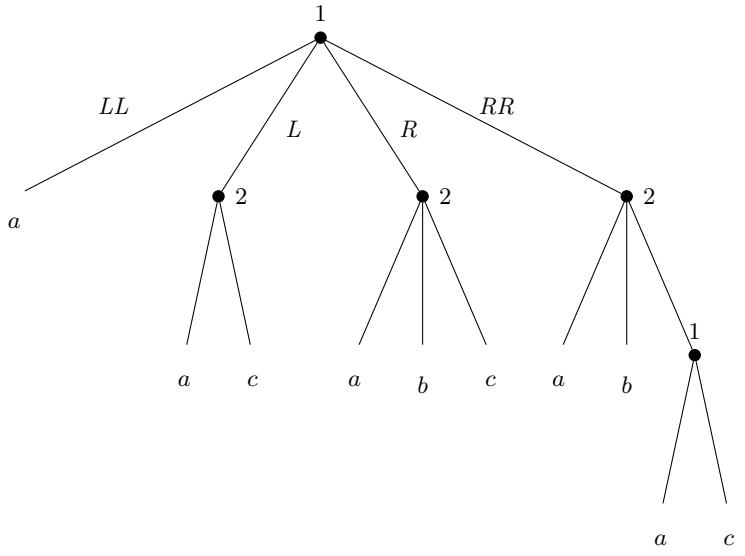

FIGURE 5. Example 4

Example 4. Here we present another example of type 2 strategically simple voting mechanism. It corresponds to the extensive form game which we show in Figure 5. Agent 1 (he) and agent 2 (she) collectively choose an outcome from three alternatives $\{a, b, c\}$. We show that agents can determine their expected utility maximizing strategies on the basis of first-order beliefs about other agents' utility functions alone.

This is obvious if an agent has a weakly dominant strategy. If agent 1 ranks a highest, he has a weakly dominant choice to choose LL at the first decision node. If agent 1 ranks b highest, he has a weakly dominant choice to choose RR at the first decision node. At the last decision node, he has a dominant choice to choose a if his preference is bac and c if his preference is bca. If agent 1 has preference cab, he has a weakly dominant choice to choose L at the first decision node. If agent 2 ranks a highest, she has a weakly dominant choice to choose a at every decision node of hers. If agent 2 ranks b highest, she has a dominant choice to choose b following actions R and RR. Following action L, she has a dominant choice to choose a if her preference is bac and c if her preference is bca. If her preference is cab, she has a weakly dominant choice to choose c following actions L and R, and to choose to offer menu $\{a, c\}$ to agent 1 following action RR.

We have thus only two cases in which there are multiple not weakly dominated strategies. In the first case, agent 1's preference is cba. For agent 1 with preference cba, LL or RR are weakly dominated at the first decision node. The choice between L and R depends only on agent 1's belief about agent 2's preference. Without loss of generality, we assume that agent 1's utility function is: $u_1(c) = 1, u_1(b) = x, u_1(a) = 0$, and that he attaches probability p (resp. q) to the event that agent 2's preference is bac (resp. bca). Agent 1 chooses L at the first decision decision node if and only if $x \leqslant \frac{q}{p+q}$.

The second case in which there are multiple undominated strategies is that agent 2 has preferences cba. She has a dominant choice to choose c following L and R. Following RR, she has a weakly dominant choice not to choose a. Whether she chooses b or chooses to offer the menu $\{a, c\}$ depends only on agent 2's belief about agent 1's preference. Without loss of generality, we assume that agent 2's utility function is: $u_2(c) = 1, u_2(b) = y, u_2(a) = 0$, and that he attaches probability $\tilde{p}$ (resp. $\tilde{q}$) to the event that agent 1's preference is bac (resp. bca). Agent 2 chooses to offer the menu $\{a, c\}$ if and only if $y \leqslant \frac{\tilde{q}}{\tilde{p}+\tilde{q}}$.

As in Example 3, there is a separation feature in the mechanism, indicated by the dashed line in Figure 6. Although agent 2's belief matters following RR, agent 1 with preference cba does not have to form a higher-order belief because, in any case, he would not choose action RR. He does not need to take into account agent 2's choice following RR. In other words, agent 2's

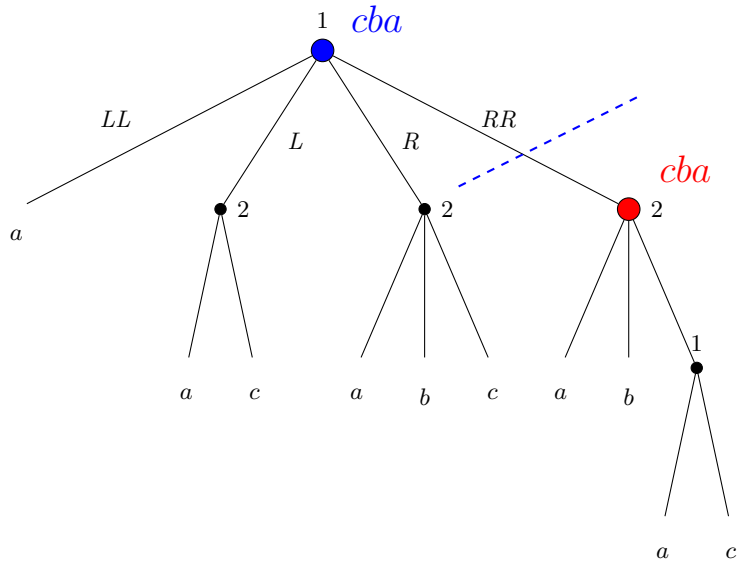

FIGURE 6. Example 4 revisited

choice following RR does not have any relevance to agent 1's choice at the first decision node if agent 1 has preference *cab*.

Figure 7 shows the reduced normal form of the extensive form game that we have shown, leaving out all strategies that are weakly dominated regardless of what agents' preferences are. Interestingly, even though at first sight the extensive form game suggests that agents have very different roles in the decision making process, the normal form game is symmetric. In Figure 8, we show the outcome correspondence implemented by the mechanism. We do not indicate the full dependence of outcomes on utility functions and beliefs. Instead, we only indicate for given ordinal preferences which outcomes are possible. If multiple outcomes are possible, then it depends on the agents' utilities and beliefs which outcome results.

The outcomes are Pareto efficient in this example, except when one agent has preference *cba* and the other agent has preference *bac*. Then it is possible that outcome a is chosen although both agents rank b higher. Agent 1 with preference *cba* chooses L at the first decision node, if he attaches sufficiently high probability to the event that agent 2 has preference *bca* (relative to *bac*). But if it turns out that agent 2 has preference *bac*, she chooses a following L.

	aaa	abb	cbb	$cc\{a,c\}$	ccb
LL	a	a	a	a	a
RRa	a	b	b	a	b
RRc	a	b	b	c	b
L	a	a	c	c	c
R	a	b	b	c	c

FIGURE 7. Normal Form for Example 4

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

FIGURE 8. Outcome Correspondence Implemented by Example 4

7. Bilateral Trade

In this section, we consider an example of an environment in which outcomes include money payments, and in which it is therefore natural to restrict attention to preferences that are monotonically increasing in money, and to beliefs that attach probability 1 to preferences that are monotonically increasing in money. The set of agents is: $I = \{S, B\}$, where S is the seller, and B is the buyer. The set of outcomes is: $A = \{\phi\} \cup T$, where "ϕ" stands for "no trade," and T is a finite subset of $\mathbb{R}_{++}$. An outcome $t \in T$ corresponds to trade at price t. We refer to any mechanism for this setting as a "bilateral trade mechanism."

The linear orders R_S over A that we consider for the seller are indexed by some value $v_S > 0$, and the linear orders R_B over A that we consider for the buyer are indexed by some value $v_B > 0$. We assume that the sets of possible values of v_i (for $i = S, B$) is a finite subset V_i of $\mathbb{R}_{++}$ with the properties: $\min V_i < \min T$, $\max V_i > \max T$, and $V_i \cap T = \varnothing$ for $i = S, B$. The linear oder with index v_S is such that the seller prefers outcome ϕ to outcome t if and only if $t < v_S$, and the seller prefers larger elements of T to smaller ones. The linear order with index v_B is such that the buyer prefers

outcome ϕ to outcome t if and only if $t > v_B$, and such that the seller prefers smaller values of T to larger ones.

In the notation of Section 5, we have now specified the sets $\mathcal{R}_i$ for $i = S, B$. The sets of admissible utility functions $\mathbf{U}_i$ and of admissible beliefs $\mathbf{M}_i$ are as given in the first sentence of Theorem 1. Note that the model that we have described does not assume quasi-linear preferences. Rather, arbitrary risk attitudes are allowed.

We assume that each agent can opt out of the mechanism; that is, each agent has a strategy that enforces the no trade outcome. Theorem 2 implies the following characterization of type 1 strategically simple bilateral trade mechanisms:

Proposition 1. *A bilateral trade mechanism is type 1 strategically simple if and only if it is equivalent to the normalform of a mechanism of the following type: Agents play a two-stage game of perfect information.*

1. *Agent i^* either chooses a price t from some finite set $\hat{T} \subseteq T$, or chooses to reject trade. If agent i^* rejects trade, then the game ends. No trade takes place, and no transfers are paid. Otherwise, Stage 2 is entered.*
2. *Agent $-i^*$ accepts or rejects trade at the price t proposed by agent i^*. If agent $-i^*$ accepts, then trade takes place, and the buyer pays the seller price t. Otherwise, no trade takes place, and no transfers are paid.*

To obtain the class of mechanisms described in Proposition 1, consider the following simple argument. When there are only two agents, the second-stage dominant strategy mechanisms as referred to in Theorem 2 are single agent mechanisms in which the agent $-i^*$ chooses among alternatives offered by agent i^*. Among the options offered that do include trade, the seller, if she is agent $-i^*$, will always pick trade at the highest price, and the buyer, if he is agent $-i^*$, will always pick trade at the lowest price. Therefore, offering trade at more than one price is redundant. Moreover, the mechanism that the seller offers must always include the no trade option.

Proposition 1 in fact provides a complete characterization of all bilateral trade mechanisms that are strategically simple, as the following result, which we prove in the Appendix, shows:

Proposition 2. *There are no bilateral trade mechanisms that are type 2 strategically simple.*

8. Related Literature

Shengwu Li [22] proposes the concept of "obviously strategy-proof mechanisms." These are a subclass of dominant strategy mechanisms in which it is particularly easy for the agents to recognize that they have a dominant strategy.[11] While Li's work is, in spirit, related to ours, our purpose is to introduce a class of mechanisms that is larger (rather than smaller) than the class of dominant strategy mechanism, yet in an interesting sense "simple." We are motivated to do so by the fact that, in many applications, the set of dominant strategy mechanisms is very small.

Li is motivated by the observation that subjects in experiments often do not recognize dominant strategies, but that they do recognize such strategies if the mechanism is "obviously strategy-proof." We are motivated by the observation that strategic reasoning that only requires agents to form first-order beliefs about other agents' utility functions, though not necessarily obvious, nevertheless seems easy.

But if subjects in experiments don't even recognize what is not obvious, the reader might object, then how can we expect them to engage in the strategic reasoning that we have called "easy" in this paper? One response is the fairly common response that in real world mechanisms the stakes are sometimes higher, and the time allowed for thinking much longer, than they are in experiments. Another response is that subjects' choices in experiments depend on the framing of the mechanism. Maybe for strategically simple mechanisms there exists some framing such that subjects engage in the reasoning that we attribute to them in this paper.

The framing might include an explanation, by the mechanism designer, of the strategic considerations that are involved when playing the mechanism. Real world mechanism designers spend a lot of time explaining to the participants in the mechanism how the mechanism works, and which considerations the participants should base their strategic choices on. It seems that in strategically simple mechanisms, the mechanism designer can present

[11] The expression "dominant strategy mechanism" that we use in this paper is synonymous with "strategy-proof mechanism."

a simple and persuasive explanation of the relevant strategic considerations to the agents. But, of course, this needs to be explored experimentally.

Some recent papers have analyzed mechanism design when agents' strategy choices are guided by "level k-thinking." The concept of "level k-thinking" is due to Stahl and Wilson [27], [28] and Nagel [24]. Crawford [12], De Clippel et. al. [15], and Kneeland [21] adapt level k-thinking to Bayesian games with incomplete information and then study mechanism design based on this concept.

There is some similarity between the agents described in our paper, who only form first-order beliefs about other agents payoffs, and level k-thinkers when $k = 2$. Level 1-thinkers don't play strictly dominated strategies, because they best-respond to an "anchor" belief. In a Bayesian game, therefore, level 2-thinkers best-respond to conjectures about other agents' strategies that are based on beliefs about these agents' types, and on the hypothesis that these types don't play undominated strategies. This is related to our agents who base their strategy choice on similar considerations.

However, there are conceptual, and technical differences between the level-k approach and our approach. The two most important are: First, level 2-thinkers stops at level 2 not because there is no value in thinking beyond level 2, but because reasoning beyond level 2 is too costly to them. In our setting, by contrast, reasoning beyond level 2 does not yield further benefits. Second, level 2-thinkers are certain about the belief anchors of other agents' types. By contrast, by allowing that agents form non-degenerate beliefs about the undominated strategies of each type of the other agents, we allow uncertainty about other agents' anchors. These differences make it hard to compare the results of authors who consider level k-thinking in mechanism design and our results.

In complete information models, De Clippel et. al. [14] and Van der Linden [29] use the number of rounds of deletion of dominated strategies, or of backward induction, that are required to solve a mechanism as a measure of the strategic complexity of mechanisms for the choice of an arbitrator or of a jury. This idea is obviously closely related to our concept of strategic simplicity. One important difference with our work is that they don't allow uncertainty about other players' preferences.

Bahel and Sprumont [1] consider dominant strategy mechanisms for the choice among Savage acts. The act that is chosen by the mechanism may

depend on each agents' beliefs about the state, but it will not depend on any agent's beliefs about the other agents' beliefs about the state, etc. This is because, for given beliefs and valuations, their mechanisms have dominant strategies. There is thus a parallel between their work and ours, although in their work beliefs are about Savage-style "states of the world," whereas in our work beliefs are about other agents' preferences.

Cremer and Riordan [13] have constructed strategically simple mechanisms for the public goods problem with quasi-linear preferences that are, in our terminology, "delegation mechanisms." Cremer and Riordan assume that the mechanism designer has some knowledge about the "delegate's" first-order belief, and that the mechanism designer can use this knowledge to appropriately design the mechanism. Cremer and Riordan then show that the designer can construct a delegation mechanism that achieves exact budget balance and ex post utilitarian optimality. It would be interesting to study strategically simple mechanisms for the public goods problem assuming less knowledge about the agents' first-order beliefs.

For certain classes of environments with quasilinear preferences, mechanisms in which agents need to form at most first-order beliefs to find their expected utility maximizing strategies have also been described in Chen and Li [11], and Yamashita and Zhu [31]. These papers also show that such mechanisms dominate the optimal dominant strategy mechanism for a revenue maximizing mechanism designer. Strategic simplicity in our sense is not the focus of these papers. But their results suggest that a further study of strategically simple mechanisms in environments with quasilinear preferences might be interesting.

9. Conclusion

Strategic simplicity as defined in this paper focuses on mechanisms in which the agents' optimal choices can be based on first-order beliefs alone. One can think of obviously strategy-proof mechanisms in the sense of [22], dominant strategy mechanisms, and our strategically simple mechanisms as successively inclusive hierarchy of mechanisms. Along this direction, it would be interesting to analyze mechanisms in which the agents' optimal choices depend on up to second-order beliefs, or finite-order beliefs. This is an important question for future research.

We have informally argued for the usefulness of the additional flexibility that strategically simple mechanisms offer over dominant strategy mechanisms. However, this paper has not made any attempt to determine among all strategically simple mechanisms the best one, for particular environments, and particular objective functions. The first obstacle to such an investigation is that a concept of "best" needs to be defined. The absence of a common prior makes this a conceptually hard problem that would be worth attacking.

Finally, our conjecture that agents' behavior in strategically simple mechanisms can be reliably predicted, once agents have been offered adequate explanations, needs to be confronted with experimental evidence. The need for the experimentalist to offer explanations of the mechanism, without "manipulating" subjects in a way that would make the experiment worthless, poses a methodological puzzle that we have not yet tackled.

References

[1] Eric Bahel and Yves Sprumont, Strategyproof Choice of Social Acts: Bilaterality, Dictatorship, and Consensuality, working paper, 2017.

[2] Salvador Barberà, Strategy-Proof Social Choice, in K. J. Arrow, A. K. Sen and K. Suzumura (eds.) *Handbook of Social Choice and Welfare* Volume 2, Netherlands: North-Holland, Chapter 25, 731-831, 2010.

[3] Pierpaolo Battigalli, Rationalizability in Infinite, Dynamic Games With Incomplete Information, *Research in Economics* 57 (2003), 1-38.

[4] Dirk Bergemann and Stephen Morris, Robust Mechanism Design, *Econometrica* 73 (2005), 1771-1813.

[5] Dmitris Bertsimas and John N. Tsitsiklis, *Introduction to Linear Optimization*, Belmont (MA): Athena Scientific, 1997.

[6] Ken Binmore, *Rational Decisions*, Princeton and Oxford: Princeton University Press, 2009.

[7] Tilman Börgers, Pure Strategy Dominance, *Econometrica* 61 (1993), 423-430.

[8] Tilman Börgers, *An Introduction to the Theory of Mechanism Design*, Oxford: Oxford University Press, 2015.

[9] Tilman Börgers and Doug Smith, Robustly Ranking Mechanisms, *American Economic Review*, Papers and Proceedings 2012, 325-329.

[10] Kalyan Chatterjee and William Samuelson, Bargaining under Incomplete Information, *Operations Research* 31 (1983), 835-851.

[11] Yi-Chun Chen and Jiangtao Li, Revisiting the Foundations of Dominant-Strategy Mechanisms, working paper, 2017.

[12] Vincent Crawford, Efficient Mechanisms for Level-k Bilateral Trading, working paper, 2016.

[13] Jacques Crémer and Michael H. Riordan, A Sequential Solution to the Public Goods Problem, *Econometrica* 53 (1985), 77-84.

[14] Geoffroy de Clippel, Kfir Eliaz, and Brian Knight, On the Selection of Arbitrators, *American Economic Review* 104 (2014), 3434-3458.

[15] Geoffroy de Clippel, Rene Saran, and Roberto Serrano, Level-k Mechanism Design, working paper, 2017.

[16] Eddie Dekel, Drew Fudenberg, and Stephen Morris, Interim Correlated Rationalizability, *Theoretical Economics* 2 (2007), 15-40.

[17] Mira Frick and Assaf Romm, Rational Behavior under Correlated Uncertainty, *Journal of Economic Theory* 160 (2015), 56-71.

[18] Allan Gibbard, Manipulation of Voting Schemes: A General Result, *Econometrica* 41 (1973), 587-601.

[19] Kathleen Hagerty and William Rogerson, Robust Trading Mechanisms, *Journal of Economic Theory* 42 (1987), 94-107.

[20] Terri Kneeland, Identifying Higher-Order Rationality, *Econometrica* 83 (2015), 2065-2079.

[21] Terri Kneeland, Mechanism Design With Level-k Types: Theory and An Application to Bilateral Trade, working paper, 2017.

[22] Shengwu Li, Obviously Strategy-Proof Mechanisms, *American Economic Review* 107 (2017), 3257-3287.

[23] Wooyoung Lim and Siyang Xiong, On Identifying Higher-Order Rationality, working paper, 2016.

[24] Rosemarie Nagel, Unraveling in Guessing Games: An Experimental Study, *American Economic Review* 85 (1995), 1313-1326.

[25] David Pearce, Rationalizable Strategic Behavior and the Problem of Perfection, *Econometrica* 52 (1984), 1029-1050.

[26] Mark Satterthwaite, Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions, *Journal of Economic Theory* 10 (1975), 187-217.

[27] Dale Stahl and Paul Wilson, Experimental Evidence on Players' Models of Other Players, *Journal of Economic Behavior and Organization* 25 (1994), 309-327.

[28] Dale Stahl and Paul Wilson, On Players' Models of Other Players: Theory and Experimental Evidence, *Games and Economic Behavior* 10 (1995), 218-254.

[29] Martin Van der Linden, Bounded Rationality and the Choice of Jury Selection Procedures, working paper, 2017.

[30] Jonathan Weinstein, The Effect of Changes in Risk Attitude on Strategic Behavior, *Econometrica* 84 (2016), 1881-1902.

[31] Takuro Yamashita and Shuguang Zhu, On the Foundations of Ex Post Incentive Compatible Mechanisms, working paper, 2017.

Appendix

Proof of Theorem 1. Sufficiency is obvious. We only prove necessity. We proceed by establishing a sequence of claims.

Claim **1.** *Let $u_i \in \mathbf{U}_i$, $u_{-i} \in \mathbf{U}_{-i}$, and let $\mu_i \in \mathbf{M}_i$ be a utility belief such that $\mu_i(\{u_{-i}\}) > 0$. Suppose $s_i, s_i' \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i)} BR_i(u_i, \hat{\mu}_i)$. Then for all $s_{-i}, s_{-i}' \in UD_{-i}(u_{-i})$:*

$$u_i(g(s_i, s_{-i})) - u_i(g(s_i', s_{-i})) = u_i(g(s_i, s_{-i}')) - u_i(g(s_i', s_{-i}')).$$

Proof of Claim *1.* Suppose the assertion were not true. Then there are $s_{-i}, s_{-i}' \in UD_{-i}(u_{-i})$ such that:

$$u_i(g(s_i, s_{-i})) - u_i(g(s_i', s_{-i})) > u_i(g(s_i, s_{-i}')) - u_i(g(s_i', s_{-i}')).$$

Pick any $\hat{\mu}_i \in \mathcal{M}_i(\mu_i)$ that places strictly positive probability on s_{-i} and s_{-i}'. Because s_i and s_i' are both in $BR_i(u_i, \hat{\mu}_i)$ both strategies must yield the same expected utility under $\hat{\mu}_i$. Now suppose we vary $\hat{\mu}_i$ such that it places ε probability more than $\hat{\mu}_i$ on s_{-i} and ε probability less than $\hat{\mu}_i$ on s_{-i}', leaving all other probabilities unchanged. If we choose $\varepsilon > 0$ and sufficiently small, we can vary $\hat{\mu}_i$ in this way so that it remains an element of $\mathcal{M}_i(\mu_i)$, and so that for the modified belief s_i is a strictly better response than s_i'. This contradicts $s_i' \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i)} BR_i(u_i, \hat{\mu}_i)$. □

Claim **2.** *Let $u_i \in \mathbf{U}_i$, $u_{-i} \in \mathbf{U}_{-i}$, and let $\mu_i, \mu_i' \in \mathbf{M}_i$ be any two utility beliefs such that $\mu_i(\{u_{-i}\}) > 0$ and $\mu_i'(\{u_{-i}\}) > 0$. Suppose*

$$s_i \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i)} BR_i(u_i, \hat{\mu}_i);$$

$$\text{and } s_i' \in \bigcap_{\hat{\mu}_i' \in \mathcal{M}_i(\mu_i')} BR_i(u_i, \hat{\mu}_i').$$

Then for all $s_{-i}, s_{-i}' \in UD_{-i}(u_{-i})$,

$$u_i(g(s_i, s_{-i})) - u_i(g(s_i', s_{-i})) = u_i(g(s_i, s_{-i}')) - u_i(g(s_i', s_{-i}')).$$

Proof of Claim *2.* We focus on the non-trivial case: $s_i \neq s_i'$. Claim 2 follows from repeated applications of Claim 1 if we can find a sequence of utility beliefs of agent i, μ_i^k $(k = 2, \ldots, K)$, and strategies of agent i, s_i^k $(k = 1, 2, \ldots, K)$, where $K \geqslant 2$, such that $s_i^1 = s_i$, $s_i^K = s_i'$, for every $k \in \{2, \ldots, K\}$ the utility belief μ_i^k places positive probability on u_{-i}, and for every $k \in \{2, \ldots, K\}$ both s_i^{k-1} and s_i^k are elements of $\bigcap_{\hat{\mu}_i^k \in \mathcal{M}_i(\mu_i^k)} BR_i(u_i, \hat{\mu}_i^k)$. We shall construct such a sequence.

For every $\alpha \in [0,1]$ we define $\mu_i(\alpha) \equiv (1-\alpha)\mu_i + \alpha\mu_i'$. We set $s_i^1 = s_i$. Define $\alpha^2 \equiv \sup\{\alpha \in [0,1] | s_i^1 \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha))} BR_i(u_i, \hat{\mu}_i)\}$. Observe that the upper hemi-continuity of the best response correspondence implies that $s_i^1 \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha^2))} BR_i(u_i, \hat{\mu}_i)\}$. If $\alpha^2 = 1$, then we can set $s_i^2 = s_i'$, $\mu_i^2 = \mu_i'$, $K = 2$, and our sequence has all the required properties.

If $\alpha^2 < 1$, define s_i^2 to be any strategy in S_i that is an element of $\bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha^2+\varepsilon))} BR_i(u_i, \hat{\mu}_i)\}$ for a sequence of $\varepsilon > 0$ tending to zero. Then, by upper hemi-continuity of the correspondence of best responses, $s_i^2 \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha^2))} BR_i(u_i, \hat{\mu}_i)\}$. We define μ_i^2 to be $\mu_i(\alpha^2)$. Note that, because μ_i and μ_i' attach strictly positive probability to u_{-i}, and because μ_i^2 is a convex combination of μ_i and μ_i', also μ_i^2 places strictly positive probability on u_{-i}. If $s_i^2 = s_i'$, then we set $K = 2$, and the construction is complete.

If $s_i^2 \neq s_i'$, then we repeat the steps just described. In general, let $k \geqslant 2$, and suppose that, after $k-1$ steps, we had determined μ_i^k such that $\mu_i^k = \mu_i(\alpha^k)$ for some $\alpha^k < 1$, and s_i^k such that $s_i^k \neq s_i'$. Then repeating the steps described above means that we define $\alpha^{k+1} \equiv \sup\{\alpha \in [\alpha^k, 1] | s_i^k \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha))} BR_i(u_i, \hat{\mu}_i)\}$. By the upper hemi-continuity of the best response correspondence: $s_i^k \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha^{k+1}))} BR_i(u_i, \hat{\mu}_i)\}$. If $\alpha^{k+1} = 1$, then we can define $s_i^{k+1} = s_i'$, $\mu_i^{k+1} = \mu_i'$, $K = k+1$, and our sequence has the required properties. If $\alpha^{k+1} < 1$, define s_i^{k+1} to be a strategy in S_i that is an element of $\bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha^{k+1}+\varepsilon))} BR_i(u_i, \hat{\mu}_i)\}$ for a sequence of $\varepsilon > 0$ tending to zero. By the upper hemicontinuity of the correspondence of best responses, $s_i^{k+1} \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i(\alpha^{k+1}))} BR_i(u_i, \hat{\mu}_i)\}$. We define μ_i^{k+1} to be $\mu_i(\alpha^{k+1})$. Note that μ_i^{k+1} places strictly positive probability on u_{-i}. If $s_i^{k+1} = s_i'$, then we set $K = k+1$, and the construction is complete. Otherwise, we continue as before.

Note that by construction, in the sequence of strategies no strategy is ever repeated. Because the number of strategies is finite, the construction has to end after a finite number of steps. At that point our sequence will have all the required properties. □

Claim **3.** *For every agent i, for every linear order $R_i \in \mathcal{R}_i$ on A, there exists a utility function u_i^* that represents R_i, such that for every $s_i \in UD_i(R_i)$ there is a strategic belief $\hat{\mu}_i$ with support equal to S_{-i} such that:*

$$BR_i(u_i^*, \hat{\mu}_i) = \{s_i\}.$$

Moreover, the utility function u_i^* can be chosen such that $u_i^*(a) - u_i^*(b) \neq u_i^*(c) - u_i^*(d)$ for all $(a,b),(c,d) \in A^2$ with $(a,b) \neq (c,d)$.

Proof of Claim 3. First note that, if we can find a utility function u_i^* with the property in the first sentence of Claim 3, then we can slightly perturb this utility function so that the property in the first sentence is maintained, but also the condition in the second sentence of Claim 3 holds. Therefore, it is sufficient to prove only the first sentence of Claim 3.

By the Lemma, and the remark in the first paragraph of the proof of that Lemma, in Börgers [7], for every strategy $s_i \in UD_i(R_i)$ there exist a utility function u_{s_i} that represents R_i, and a full support strategic belief $\hat{\mu}_i$, such that s_i is the unique maximizer of expected utility given that belief. To prove Claim 3 it therefore only remains to be shown that the utility functions u_{s_i} can be chosen to be the same for all strategies $s_i \in UD_i(R_i)$.

We begin with the following observation: Suppose that s_i is the unique maximizer of expected utility in S_i for utility function u_i and full support strategic belief $\hat{\mu}_i$, and suppose that $f : \mathbb{R} \to \mathbb{R}$ is strictly increasing and concave. We claim that then there is another full support strategic belief $\hat{\hat{\mu}}_i$ such that s_i is the unique maximizer of expected utility for the utility function $f \circ u_i$. To see this note first that, because s_i maximizes expected utility for a full support belief if utility is u_i, it is not weakly dominated given utility function u_i. Next, because f is increasing and concave, s_i is not weakly dominated given utility function $f \circ u_i$, either. This follows directly from the argument in the proof of Proposition 1 in Weinstein [30]. We can now use Lemma 4 in Pearce [25] and conclude that there is some full support strategic belief $\hat{\hat{\mu}}_i$ of agent i such that s_i maximizes expected utility when the utility function is $f \circ u_i$. It remains to be shown that this belief can be chosen such that s_i is the *unique* maximizer of expected utility. We do this in the next paragraph.

Because s_i is the unique maximizer of expected utility for some full support belief if the utility function is u_i, by Theorem 2.3 in Bertsimas and Tsitsiklis [5], the utility vector $(u_i(s_i, s_{-i}))_{s_{-i} \in S_{-i}} \in \mathbb{R}^{|S_{-i}|}$ is an extreme point of the convex hull of the set of all such utility vectors:

$$co\left(\left\{\left(u_i(s_i', s_{-i})\right)_{s_{-i} \in S_{-i}} | s_i' \in S_i\right\}\right).$$

We now claim that the utility vector corresponding to s_i remains an extreme point if we apply an increasing and concave transformation to u_i. That is,

we claim that $(f(u_i(s_i, s_{-i})))_{s_{-i}\in S_{-i}} \in \mathbb{R}^{|S_{-i}|}$ is an extreme point of:

$$co\left(\left\{\left(f(u_i(s_i', s_{-i}))\right)_{s_{-i}\in S_{-i}} | s_i' \in S_i\right\}\right).$$

Suppose it were not. Then $(f(u_i(s_i, s_{-i})))_{s_{-i}\in S_{-i}}$ could be written as a convex combination of the elements of $\left\{(f(u_i(s_i', s_{-i})))_{s_{-i}\in S_{-i}} | s_i' \in S_i, s_i' \neq s_i\right\}$, that is, there would be a mixed strategy $\sigma_i \in \Delta(S_i)$ of agent i that attaches zero probability to s_i, and such that:

$$(f(u_i(s_i, s_{-i})))_{s_{-i}\in S_{-i}} = \sum_{s_i'\in S_i} \left(f(u_i(s_i', s_{-i}))\right)_{s_{-i}\in S_{-i}} \sigma_i(s_i').$$

Because f is strictly concave, this implies:

$$(u_i(s_i, s_{-i}))_{s_{-i}\in S_{-i}} \lneqq (u_i(\sigma_i, s_{-i}))_{s_{-i}\in S_{-i}},$$

which contradicts that s_i is not weakly dominated for utility function u_i. We conclude that $(f(u_i(s_i, s_{-i})))_{s_{-i}\in S_{-i}} \in \mathbb{R}^{|S_{-i}|}$ is an extreme point. Using again Theorem 2.3 in Bertsimas and Tsitsiklis [5] this implies that there is some function $\xi : S_{-i} \to \mathbb{R}$ such that s_i is the unique maximizer of $\sum_{s_{-i}\in S_{-i}} \xi(s_{-i}) f(u_i(s_i, s_{-i}))$ in S_i. Let us treat ξ as a vector in $\mathbb{R}^{|S_{-i}|}$. One can verify that there must be a small ball around ξ such that for every vector $\tilde{\xi}$ in this ball s_i is the unique maximizer of $\sum_{s_{-i}\in S_{-i}} \tilde{\xi}(s_{-i}) f(u_i(s_i, s_{-i}))$. We can pick from this ball some $\tilde{\xi}$ such that $\sum_{s_{-i}\in S_{-i}} \tilde{\xi}(s_{-i}) \neq 0$. Now consider the vector $\tilde{\mu}_i$ defined by:

$$\tilde{\mu}_i(s_{-i}) \equiv \frac{\hat{\hat{\mu}}_i + \varepsilon \frac{\tilde{\xi}(s_{-i})}{\sum_{s'_{-i}\in S_{-i}} \tilde{\xi}(s'_{-i})}}{1+\varepsilon}$$

for all $s_{-i} \in S_{-i}$. For sufficiently small $\varepsilon > 0$ this is a strategic belief. It is a convex combination of $\hat{\hat{\mu}}_i$, for which s_i is *a* expected utility maximizer, and of $\tilde{\xi}$, for which s_i is the *unique* maximizer of $\sum_{s_{-i}\in S_{-i}} \tilde{\xi}(s_{-i}) f(u_i(s_i, s_{-i}))$ in S_i. Therefore, s_i is the unique expected utility maximizer for the strategic belief $\tilde{\mu}_i$.

We can now complete the proof by showing that there are a utility function u_i^* and, for every $s_i \in UD_i(R_i)$, a concave function $f_{s_i} : \mathbb{R} \to \mathbb{R}$, such that $u_i^* = f_{s_i}(u_{s_i})$ for all $s_i \in UD_i(R_i)$. We first construct u_i^*. Enumerate the elements of A as $a_1, a_2, \ldots, a_L$ such that $a_L R_i a_{L-1} R_i a_{L-2} R_i \ldots R_i a_1$. We

pick u_i^* to satisfy the following, where the first two lines are a normalization:

$$\begin{aligned} u_i^*(a_1) &= 0 \\ u_i^*(a_2) &= 1 \\ &\dots \\ u_i^*(a_{\ell-1}) < u_i^*(a_\ell) &< u_i^*(a_{\ell-1}) + \dots \\ \dots (u_i^*(a_{\ell-1}) - u_i^*(a_{\ell-2})) &\min_{s_i \in UD_i(R_i)} \frac{u_{s_i}(a_\ell) - u_{s_i}(a_{\ell-1})}{u_{s_i}(a_{\ell-1}) - u_{s_i}(a_{\ell-2})}. \end{aligned}$$

Note that the right most term in the inequality is strictly larger than the left term, so that u_i^* can be constructed, and will be monotonically increasing, and thus compatible with R_i.

We now turn to the construction of the functions f_{s_i}. For every s_i we set $f_{s_i}(u_{s_i}(a_\ell)) = u_i^*(a_\ell)$ for all $\ell = 1, 2, \dots, L$. This defines f_{s_i} for a finite number of elements of $\mathbb{R}$ only. However, it is clear that we can extend f_{s_i} to a concave piecewise linear function on $\mathbb{R}$ if it satisfies the following concavity condition for the points in which it is defined:

$$\frac{f_{s_i}(u_{s_i}(a_\ell)) - f_{s_i}(u_{s_i}(a_{\ell-1}))}{u_{s_i}(a_\ell) - u_{s_i}(a_{\ell-1})} \leqslant \frac{f_{s_i}(u_{s_i}(a_{\ell-1})) - f_{s_i}(u_{s_i}(a_{\ell-2}))}{u_{s_i}(a_{\ell-1}) - u_{s_i}(a_{\ell-2})}$$

for all $\ell \geqslant 2$. By the definition of f_{s_i}, this inequality is equivalent to:

$$\begin{aligned} \frac{u_i^*(a_\ell) - u_i^*(a_{\ell-1})}{u_{s_i}(a_\ell) - u_{s_i}(a_{\ell-1})} \leqslant \frac{u_i^*(a_{\ell-1}) - u_i^*(a_{\ell-2})}{u_{s_i}(a_{\ell-1}) - u_{s_i}(a_{\ell-2})} &\Leftrightarrow \\ u_i^*(a_\ell) \leqslant u_i^*(a_{\ell-1}) &+ \dots \\ \dots (u_i^*(a_{\ell-1}) - u_i^*(a_{\ell-2})) &\frac{u_{s_i}(a_\ell) - u_{s_i}(a_{\ell-1})}{u_{s_i}(a_{\ell-1}) - u_{s_i}(a_{\ell-2})} \end{aligned}$$

which holds by construction. □

Claim **4.** *For every agent i, for every linear order $R_i \in \mathcal{R}_i$ on A, and for every $u_{-i} \in \mathbf{U}_{-i}$ either*

(i) there is for every strategy $s_i \in UD(R_i)$ an alternative a such that $g(s_i, s_{-i}) = a$ for all $s_{-i} \in UD_{-i}(u_{-i})$,

or

(ii) there is for every strategy combination $s_{-i} \in UD_{-i}(u_{-i})$ an alternative a such that $g(s_i, s_{-i}) = a$ for all $s_i \in UD_i(R_i)$,

or both.

Proof of Claim 4. Let us represent R_i by the utility function u_i^* from Claim 3. Pick any two $s_i, s_i' \in UD_i(R_i)$. By Claim 3 there are a full support strategic belief $\hat{\mu}_i$ such that: $BR_i(u_i^*, \hat{\mu}_i) = \{s_i\}$, and a full support strategic belief $\hat{\mu}_i'$ such that: $BR_i(u_i^*, \hat{\mu}_i') = \{s_i\}$. Because $\hat{\mu}_i$ has full support, and because every strategy is undominated for at least some utility function, there is a utility belief μ_i with $\mu_i(u_{-i}) > 0$ that is compatible with $\hat{\mu}_i$. Similarly, there is a utility belief μ_i' with $\mu_i'(u_{-i}) > 0$ that is compatible with $\hat{\mu}_i'$. This implies $s_i \in \bigcap_{\hat{\mu}_i \in \mathcal{M}_i(\mu_i)} BR_i(u_i, \hat{\mu}_i)$ and $s_i' \in \bigcap_{\hat{\mu}_i' \in \mathcal{M}_i(\mu_i')} BR_i(u_i, \hat{\mu}_i')$. Therefore, by Claim 2 for all $s_{-i}, s_{-i}' \in UD_{-i}(u_{-i})$:

$$u_i^*(g(s_i, s_{-i})) - u_i^*(g(s_i', s_{-i})) = u_i^*(g(s_i, s_{-i}')) - u_i^*(g(s_i', s_{-i}')). \qquad (*)$$

This has to hold for any two $s_i, s_i' \in UD_i(R_i)$.

Now let us fix some $s_i \in UD_i(R_i)$, and suppose first that for some $a \in A$ we have: $g(s_i, s_{-i}) = a$ for all $s_{-i} \in UD_{-i}(u_{-i})$. Then (*) implies that for every other $s_i' \in UD_i(R_i)$ there must be some $\tilde{a} \in A$ such that $g(s_i, s_{-i}) = \tilde{a}$ for all $s_{-i} \in UD_{-i}(u_{-i})$. This follows from $u_i^*(a) - u_i^*(b) \neq u_i^*(c) - u_i^*(d)$ for all $(a, b), (c, d) \in A^2$ with $(a, b) \neq (c, d)$. Thus, we have obtained Case (i).

Next suppose that for the s_i that we fixed in the previous paragraph we have: $g(s_i, s_{-i}) \neq g(s_i, s_{-i}')$ for some $s_{-i}, s_{-i}' \in UD_{-i}(u_{-i})$. Then $u_i^*(a) - u_i^*(b) \neq u_i^*(c) - u_i^*(d)$ for all $(a, b), (c, d) \in A^2$ with $(a, b) \neq (c, d)$ implies that (*) can only hold if both sides equal zero, and hence $g(s_i, s_{-i}) = g(s_i', s_{-i})$ for all $s_i, s_i' \in UD_i(R_i)$ and all $s_{-i} \in UD_{-i}(R_{-i})$. Thus, we have obtained Case (ii). □

Claim **5.** *Suppose for every agent j we have a linear order $R_j \in \mathcal{R}_j$ on A. Then, for every agent i, either*

(i) there is for every strategy $s_i \in UD(R_i)$ an alternative a such that $g(s_i, s_{-i}) = a$ for all $s_{-i} \in UD_{-i}(R_{-i})$,

or

(ii) there is for every strategy combination $s_{-i} \in UD_{-i}(R_{-i})$ an alternative a such that $g(s_i, s_{-i}) = a$ for all $s_i \in UD_i(R_i)$,

or both.

Proof of Claim 5. Claim 5 follows from Claim 4 if we represent for each j with $j \neq i$ the linear order R_j by the utility function u_j^* referred to in Claim 3 because then: $UD_{-i}(u_{-i}^*) = UD_{-i}(R_{-i})$. □

Completing the Proof of Theorem 1: The claim is obviously true if there is an alternative a such that $g(s) = a$ for all $s \in UD(R)$. Therefore

from now on we restrict attention in this proof to the case that there are two alternatives $a \neq b$ such that $g(s) = a$ for some $s \in UD(R)$ and $g(s') = b$ for some other $s' \in UD(R)$.

We shall say that agent $i \in I$ "has no influence" if for every $s_{-i} \in UD_{-i}(R_{-i})$ there is an $a \in A$ such that $g(s_i, s_{-i}) = a$ for all $s_i \in UD_i(R_i)$, and we shall say that agent i is a dictator if agent i has the property ascribed to agent i^* in Theorem 1. By Claim 5 every agent i either has no influence, or is a dictator.

Next note that it cannot be that there is more than one dictator. A dictator can enforce any of the alternatives contained in $\{g(s)|s \in UD(R)\}$. We have assumed that there are at least two such alternatives, say a and b. Having two dictators leads to a contradiction if one of them chooses an action that enforces a, and the other one chooses an action that enforces b.

Finally note that it cannot be that all agents have no influence. Recall that we are considering the case in which there are two alternatives $a \neq b$ such that $g(s) = a$ for some $s \in UD(R)$ and $g(s') = b$ for some other $s' \in UD(R)$. Consider the sequence of n strategy combinations s^k obtained by switching sequentially first agent 1, then agent 2, etc. from strategy s_i to strategy s'_i. Thus, $s^1 = (s'_1, s_2, \ldots, s_n)$, $s^2 = (s'_1, s'_2, s_3 \ldots, s_n)$, etc. Define $s^0 = s$. Because $g(s^0) \neq g(s^n)$, there must be some k such that $g(s^k) \neq g(s^{k-1})$. But this means that by construction agent k has influence. Hence agent k must be a dictator. □

Proof of Theorem 2, Part (ii). Consider a type 1 strategically simple mechanism, and let

$$i^* \in \bigcap_{R \in \times_{i \in I} \mathcal{R}_i} I^*(R).$$

We shall show that, for all $i \neq i^*$ and all $R_i \in \mathcal{R}_i$, the set $UD_i(R_i)$ contains exactly one element. Suppose that, for some i and R_i, the set $UD_i(R_i)$ had two distinct elements, say s_i and s'_i. Consider any $s_{-i} \in S_{-i}$. We claim that $g(s_i, s_{-i}) = g(s'_i, s_{-i})$. To see this, first note that $s_{-i} \in UD_{-i}(R_{-i})$ for some $R_{-i} \in \times_{j \neq i} \mathcal{R}_j$, because we assume that every strategy is not weakly dominated for some utility function. Now consider the preference profile (R_i, R_{-i}). Since agent i^* is local dictator for this preference profile, for any $s^*_i \in UD_{i*}(R_{i*})$, there is an $a \in A$ such that: $g(s_{i*}, s_{-i*}) = a$ for all $s_{-i*} \in UD_{-i*}(R_{-i*})$. This implies: $g(s_{i*}, s_i, s_{-(i*,i)}) = g(s_{i*}, s'_i, s_{-(i*,i)})$ for all $s_{-(i*,i)} \in UD_{-(i*,i)}(R_{-(i*,i)})$. As this holds for all $s^*_i \in UD_{i*}$, the assertion

follows. But this contradicts our assumption that mechanisms do not have duplicate strategies.

Fix any $s_{i*} \in S_{i*}$, and consider the mechanism in which we have removed agent i^* from the set of agents, in which all other agents have the same strategy sets as originally, i.e., S_j, and in which the outcome corresponding to any s_{-i*} is given by $g(s_{i*}, s_{-i*})$. Let us call this mechanism the "restricted mechanism" corresponding to s_{i*}. If all agents $j \neq i$ play the strategies that are uniquely dominant in the overall mechanism, then the restricted mechanism implements an outcome function: $F_{s_{i*}} : \mathbf{U}_{-i} \times \mathbf{M}_{-i} \to A$. Because, in the overall mechanism, agents have dominant strategies, the outcome correspondence is constant with respect to beliefs, and it is also constant if utility functions are changed without changing the order of the elements of A. We can therefore write F as: $F_{s_{i*}} : \times_{j \neq i^*} \mathcal{R}_j \to A$. We can treat this outcome function as a direct mechanism. Because agents choose dominant strategies in the overall mechanism, in the direct mechanism it is a dominant strategy for each agent to report their preferences truthfully.

Because in the overall mechanism agents have uniquely dominant strategies, they must have for every preference ordering a strategy that induces in each of the restricted mechanisms a dominant strategy. Agent i^* thus expects, for each of the strategies that he can choose, the same outcome distribution as he would in the sequential mechanism described in Theorem 2, if the second stage mechanisms are the restricted mechanisms described by the outcome function $F_{s_{i*}}$. Agent i^* will make the same choice as in the sequential mechanism as in the given type 1 strategically simple mechanism. This implies part (ii) of Theorem 2. □

Proof of Proposition 2. Before we focus on the bilateral trade problem, we prove the following lemma that is true in the general framework introduced in Section 2, and not just in the bilateral trade framework.

Lemma 1. *Let $i \in I$, $R_i, \hat{R}_i \in \mathcal{R}_i$, and $s_i \in UD_i(R_i)$. Then there exists an $\hat{s}_i \in UD_i(\hat{R}_i)$ such that for any $s_{-i} \in S_{-i}$, $g(\hat{s}_i, s_{-i}) \hat{R}_i g(s_i, s_{-i})$.*

Proof of Lemma 1. Either $s_i \in UD_i(\hat{R}_i)$, in which case the lemma is true if we set $\hat{s}_i = s_i$, or $s_i \notin UD_i(\hat{R}_i)$, in which case, because we are considering finite mechanisms, there is some $s_i' \in UD_i(\hat{R}_i)$ that weakly dominates s_i, and the lemma follows if we set $\hat{s}_i = s_i'$. □

Now we turn to the bilateral trade problem. To simplify the notation, we shall use "v_i" not just to refer to agent i's value of the object, but also to refer to the corresponding ordinal preference. We use $UD_i(v_i)$ to denote the set of strategies of agent i that are not weakly dominated if agent i has ordinal preference v_i. We use $I^*(v_S, v_B)$ to denote the set of local dictators at preference profile (v_S, v_B). Finally, for any (v_S, v_B), we denote by $\mathcal{O}(v_S, v_B)$ the set of outcomes that can arise when both agents play strategies that are not weakly dominated given their valuations. That is, $\mathcal{O}(v_S, v_B) \equiv \{a \in A \,|\, a = g(s_S, s_B)$ for some $s_S \in UD_S(v_S)$ and $s_B \in UD_B(v_B)\}$.

We now prove four claims that will be useful in the proof of the proposition. These claims describe implications of strategic simplicity in the bilateral trade setting, regardless of whether we are referring to type 1 or type 2 strategic simplicity.

CLAIM **1.** *If $I^*(v_S, v_B) = \{S, B\}$, then $|\mathcal{O}(v_S, v_B)| = 1$.*

Proof of CLAIM *1.* This immediately follows from the definition of local dictatorship: if one agent were able to enforce two different outcomes, then the other agent could not be a local dictator. □

CLAIM **2.** *If $I^*(v_S, v_B) = \{i^*\}$ for some $i^* \in I$, then $|\mathcal{O}(v_S, v_B)| \geqslant 2$, and $\mathcal{O}(v_S, v_B) \cap T \neq \varnothing$.*

Proof of CLAIM *2.* The first part of the claim follows from the fact that if $\mathcal{O}(v_S, v_B)$ had just one element, then both agents would be local dictators. The second part of the claim is a direct implication of the first part. □

CLAIM 2 implies that the following notations for pairs (v_S, v_B) such that $I^*(v_S, v_B) = \{i^*\}$ for some $i^* \in I$ are well-defined: $\bar{t}(v_S, v_B) \equiv \max \mathcal{O}(v_S, v_B) \cap T$ and $\underline{t}(v_S, v_B) \equiv \min \mathcal{O}(v_S, v_B) \cap T$.

Next, we show that the assumption that each agent has an opting out strategy implies that only ex post individually rational outcomes can occur when agents do not choose weakly dominated strategies.

CLAIM **3.** *For any $(v_S, v_B) \in V_S \times V_B$, for every $i \in \{S, B\}$, agent i with preferences v_i weakly prefers every outcome in $\mathcal{O}(v_S, v_B)$ to no trade.*

Proof of CLAIM *3.* The claim is straightforward for outcomes when both agents are local dictators. By Lemma 1, each agent weakly prefers at least one outcome in $\mathcal{O}(v_S, v_B)$ to no trade. By Claim 1, if both agents are local

dictators, $\mathcal{O}(v_S, v_B)$ has just one element. Hence, both agents must weakly prefer this outcome to no trade.

In the rest of the proof we focus on the case of a unique local dictator, $I(v_S, v_B) = \{i^*\}$. Consider first the agent who is *not* the local dictator, i.e. agent $i \neq i^*$. Obviously, it is sufficient to consider only outcomes in $\mathcal{O}(v_S, v_B)$ that correspond to trade at some price $t \in T$. Consider any strategy $s_{i*} \in UD_{i*}(v_{i*})$ of agent i^* that results in trade at price t against any strategy in $UD_i(v_i)$ (see Figure 9). Because i has an opting out strategy, by Lemma 1 one of the strategies in $UD_i(v_i)$ must yield at least as good an outcome as no trade for agent i with preferences v_i. This implies that trade at price t must be at least as good as no trade for agent i with preference v_i.

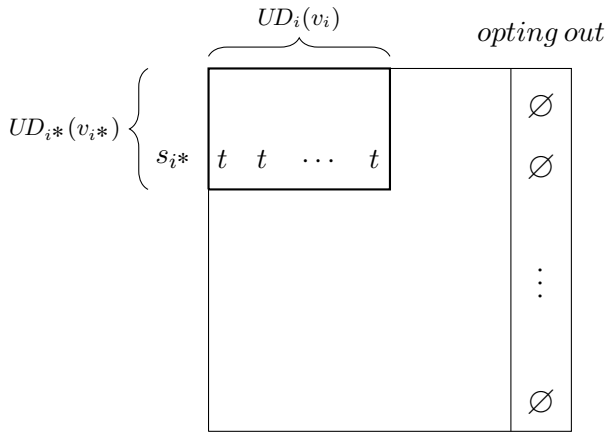

FIGURE 9. Agent i^* is the unique local dictator at (v_S, v_B). In this case, trade at price t must be at least as good as no trade for agent i with preference v_i.

We are left with the task to show that, when there is a unique local dictator i^*, all outcomes are ex post individually rational for the local dictator herself. Without loss of generality we consider the case $i^* = S$. Our proof strategy will be the following. We consider any (v_S, v_B) such that ex post individual rationality for the seller is violated at (v_S, v_B). We show that then there must be some $v'_S > v_S$ and some v'_B such that $I^*(v'_S, v'_B) = \{S\}$ and ex post individual rationality for the seller is also violated at (v'_S, v'_B). This implies the claim, because the assumption that there is any value profile at which the seller's ex post individual rationality were violated would

imply that there would have to be a largest $v_S \in I_S^*$ for which individual rationality is violated for some v_B, and this would be in contradiction with the assertion that we just made.

Thus, consider any (v_S, v_B) such that $I^*(v_S, v_B) = \{S\}$ and the seller's individual rationality is violated at (v_S, v_B) (see Figure 10). This means that there is a strategy $s_S \in UD_S(v_S)$ for which $g(s_S, s_B) \in T$, and $g(s_S, s_B) < v_S$ for all $s_B \in UD_B(v_B)$. To start, note that there must exist some $v'_B \in V_B$ and $s'_B \in UD_B(v'_B)$ such that v_S ranks $g(s_S, s'_B)$ above no trade. Otherwise, for the seller with preference v_S, the strategy s_S would be weakly dominated by the strategy of opting out. Since v_S ranks $g(s_S, s'_B)$ above no trade, we have $g(s_S, s'_B) \in T$ and $g(s_S, s'_B) > v_S$.

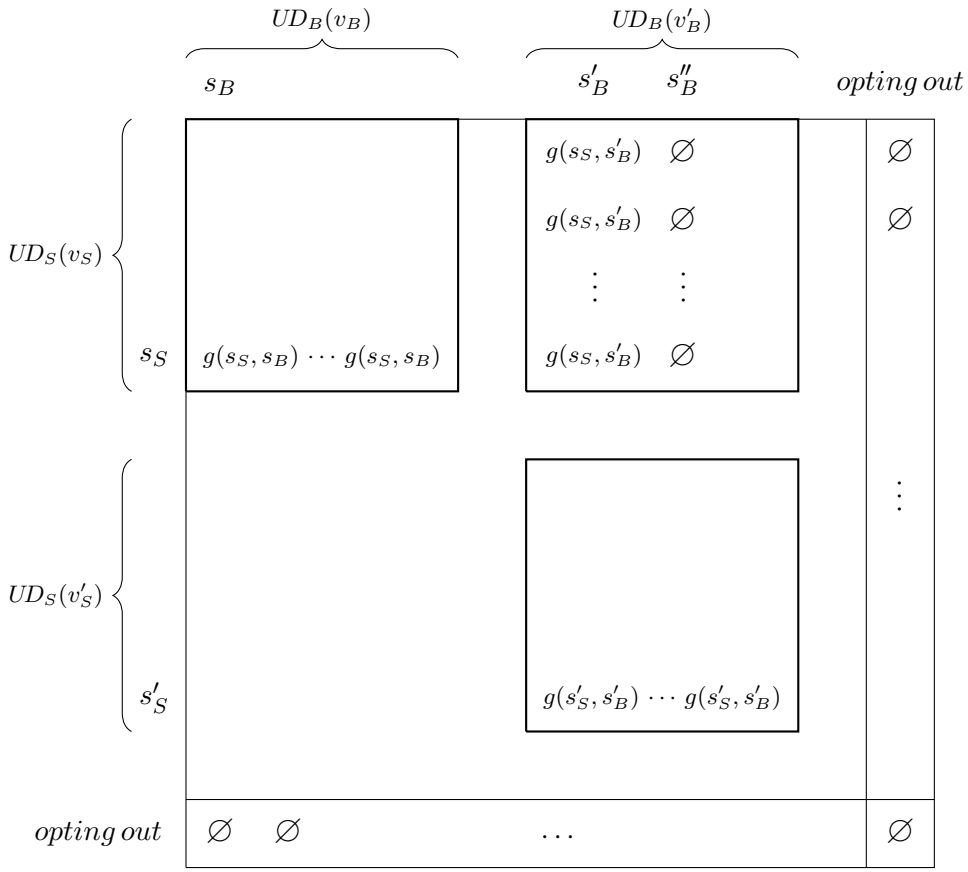

FIGURE 10. The seller is the unique local dictator at (v_S, v_B). Suppose that $g(s_S, s_B) < v_S$, we find another $v'_S > v_S$ such that $g(s'_S, s'_B) < v'_S$.

Our next objective is to prove the following statements about the behavior of the mechanism at (v_S, v_B) and (v_S, v'_B). Here, s_B is any arbitrary strategy in $UD(v_B)$.

(i) B is the unique local dictator at (v_S, v'_B);
(ii) $g(s_S, s'_B) > g(s_S, s_B)$;
(iii) $g(s_S, s_B) > v'_B$;
(iv) $g(s_S, s'_B) > v'_B$.

Proving (ii) is simple: We have $g(s_S, s_B) < v_S$, and, by construction, $g(s_S, s'_B) > v_S$. Thus, (ii) follows. Now note that $g(s_S, s'_B) > g(s_S, s_B)$ implies that v'_B ranks $g(s_B, s'_B)$ below $g(s_S, s_B)$. By Lemma 1, there must be some strategy $s''_B \in UD_B(v'_B)$ such that v'_B ranks $g(s_S, s''_B)$ above $g(s_S, s_B)$ or $g(s_S, s''_B) = g(s_S, s_B)$. Note that we can conclude $g(s_S, s_B) \neq g(s_S, s''_B)$, and hence that (i) is true.

As an intermediate step we show next that $g(s_S, s''_B) = \phi$. If $g(s_S, s''_B)$ were an element of T, since v'_B ranks $g(s_S, s''_B)$ above $g(s_S, s_B)$ or $g(s_S, s''_B) = g(s_S, s_B)$, it would have to be that $g(s_S, s''_B) \leqslant g(s_S, s_B)$. Since v_S ranks $g(v_S, v_B)$ below no trade, v_S also ranks $g(s_S, s''_B)$ below no trade. But this contradicts the ex post individual rationality for the agent who is not dictator, which we showed in an earlier step of this proof. We conclude: $g(s_S, s'_B) = \phi$.

By construction, v'_B ranks $g(s_S, s''_B)$ above $g(s_S, s_B)$, hence (iii) follows from the fact that $g(s_S, s''_B)$ is no trade. Finally, (ii) and (iii) imply (iv).

Now note that we have obtained a pair of valuations at which the buyer is the local dictator, and, by (iv), the buyer's ex post individual rationality is violated. We can therefore repeat the argument just presented, reversing the roles of the buyer and the seller. This yields the conclusion that there is some $v'_S \in V_S$, and some $s'_S \in UD(v'_S)$ such that $g(s'_S, s'_B) \in T$ and $g(s'_S, s'_B) < v'_B$, and:

(v) S is the local dictator at (v'_S, v'_B);
(vi) $g(s'_S, s'_B) < g(s_S, s'_B)$;
(vii) $g(s_S, s'_B) < v'_S$;
(viii) $g(s'_S, s'_B) < v'_S$.

The proof can now be concluded. By construction: $v_S < g(s_S, s'_B)$. Result (vii) says: $g(s_S, s'_B) < v'_S$. Hence $v_S < v'_S$. Moreover $(viii)$ shows that ex post individual rationality for the seller is violated at (v'_S, v'_B). □

Our next result shows that, if at some valuation profile some agent i is the unique local dictator, this agent remains (not necessarily unique) local dictator even if we change i's valuation, keeping the other valuation fixed.

CLAIM **4.** *Suppose* $I^*(v_i, v_{-i}) = \{i\}$. *Then* $i \in I^*(v_i', v_{-i})$ *for all* $v_i' \in V_i$.

Proof of CLAIM *4.* Without loss of generality we focus on the case $i = S$. The proof is indirect. Let $I^*(v_S, v_B) = \{S\}$, and suppose $I^*(v_S, v_B) = \{B\}$ for some $v_S' \in V_S$. Let $s_S \in S_S$ be the strategy in $UD_S(v_S)$ that enforces the outcome $\bar{t}(v_S, v_B)$ against any strategy in $UD_B(v_B)$. Let $s_B \in S_B$ be the strategy in $UD_B(v_B)$ that enforces the outcome $\bar{t}(v_S', v_B)$ against any strategy in $UD_S(v_S')$.

Suppose also, first, that: $\bar{t}(v_S', v_B) > \bar{t}(v_S, v_B)$. By Lemma 3, v_S ranks $\bar{t}(v_S, v_B)$ above no trade. Therefore, v_S must also rank $\bar{t}(v_S', v_B)$ above no trade. By Lemma 1, the seller with value v_S must have a strategy in $UD_S(v_S)$ that guarantees an outcome at least as good as $\bar{t}(v_S', v_B)$ against any strategy in $UD_B(v_B)$. This contradicts the definition of $\bar{t}(v_S, v_B)$ as the highest price that the seller can guarantee with a strategy in $UD_S(v_S)$.

Now suppose: $\bar{t}(v_S', v_B) < \bar{t}(v_S, v_B)$. By Lemma 1, the seller with value v_S' must have at least one strategy in $UD_S(v_S')$ that yields against s_B an outcome at least as good as $\bar{t}(v_S, v_B)$. This contradicts that s_B yields $\bar{t}(v_S', v_B)$ for all $s_S \in UD_S(v_S)$.

Finally suppose: $\bar{t}(v_S', v_B) = \bar{t}(v_S, v_B)$. Let $s_B' \in UD_B(v_B)$ denote a strategy such that $g(s_S, s_B') \neq \bar{t}(v_S', v_B)$ for all $s_S \in UD_S(v_S')$. By Claim 2, such an s_B' exists. Since the buyer is the unique local dictator at preference profile (v_S', v_B), by Lemma 1, any strategy in $UD_S(v_S')$ yields against s_B' an outcome at least as good as $\bar{t}(v_S, v_B)$. This outcome cannot be trade at price $\bar{t}(v_S, v_B)$, because s_B' leads to an outcome other than $\bar{t}(v_S, v_B)$, and it cannot be trade at a price higher than $\bar{t}(v_S, v_B)$ because we are considering the case $\bar{t}(v_S', v_B) = \bar{t}(v_S, v_B)$. Therefore, any strategy in $UD_S(v_S')$ yields against s_B' no trade. But then we have concluded that the seller prefers no trade to trade at $\bar{t}(v_S, v_B)$, which contradicts Claim 3, i.e. the seller's ex post individual rationality at (v_S, v_B) and at (v_S', v_B). □

We now turn to an indirect proof of Proposition 2, that is, we postulate that a bilateral trade mechanism is type 2 strategically simple, and then derive a contradiction. The next four claims describe implications of the premises of the indirect proof.

CLAIM **5.** *There are $v_S, \hat{v}_S \in V_S$ with $v_S \neq \hat{v}_S$ and $v_B, \hat{v}_B \in V_B$ with $v_B \neq \hat{v}_B$ such that: $I^*(v_S, v_B) = \{S\}, I^*(\hat{v}_S, \hat{v}_B) = \{B\}$, and $I^*(v_S, \hat{v}_B) = I^*(\hat{v}_S, v_B) = \{S, B\}$.*

Proof of CLAIM *5.* By definition of type 2 strategic simplicity, we must have two pairs of values in $V_S \times V_B$, one at which S is the unique local dictator, and another one at which B is the unique local dictator. By Claim 4 these two pairs must have no component in common. Claim 4 also implies that if we combine the seller's value in one pair with a buyer's value in the other pair, then both agents must be local dictators. □

For the remainder of the proof we use the notation (v_S, v_B) and $(\hat{v}_S, \hat{v}_B)$ to refer to the two pairs the existence of which is asserted in Claim 5.

CLAIM **6.** $\mathcal{O}(v_S, \hat{v}_B) = \{\phi\}$.

Proof of CLAIM *6.* By Claim 1, $\mathcal{O}(v_S, \hat{v}_B)$ has only one element. Suppose $\mathcal{O}(v_S, \hat{v}_B) = \{t\}$ for some $t \in T$. Using Lemma 1 for the buyer, we can infer $t \leqslant \underline{t}(v_S, v_B)$. Because at (v_B, v_S) the seller is the only local dictator, Claim 2 implies that the set $\mathcal{O}(v_S, v_B)$ must include an outcome a other than $\underline{t}(v_S, v_B)$. If this is trade at a price higher than $\underline{t}(v_S, v_B)$, then clearly the buyer strictly prefers $\underline{t}(v_S, v_B)$ to a. But if a is no trade, then Claim 3 implies that the buyer strictly prefers $\underline{t}(v_S, v_B)$ to a. Thus, $\mathcal{O}(v_S, v_B)$ includes an outcome a that the buyer ranks strictly below $\underline{t}(v_S, v_B)$, and hence also strictly below t. The seller has a strategy that locally enforces this outcome at (v_S, v_B). By Lemma 1 this contradicts the fact that the buyer has a strategy that enforces at $(v_S, \hat{v}_B)$ the price t. □

CLAIM **7.** *$v_S > \hat{v}_S$ and $v_B > \hat{v}_B$.*

Proof of CLAIM *7.* The arguments are symmetric for seller and buyer. Consider the seller. Because no trade occurs at $(v_S, \hat{v}_B)$, by Claim 6, and trade at some price t is a possible outcome at $(\hat{v}_S, \hat{v}_B)$, Lemma 1 implies that with value v_S the seller must find no trade preferable to a trade at price t. Claim 3 says that the seller with value $\hat{v}_S$ prefers trade at price t to no trade. These findings together imply $v_S > \hat{v}_S$. □

CLAIM **8.** *$\mathcal{O}(\hat{v}_S, v_B) = \{t^*\}$ for some $t^* \in T$.*

Proof of CLAIM *8.* By Claim 1, $\mathcal{O}(\hat{v}_S, v_B)$ has only one element. Suppose $\mathcal{O}(\hat{v}_S, v_B) = \{\phi\}$. By Claims 2 and 3, trade at some price is contained in $\mathcal{O}(v_S, v_B)$ that the seller with value v_S strictly prefers to no trade. When

the seller has value $\hat{v}_S$, the seller still strictly prefers trade at that price to no trade, because, by Claim 7, $\hat{v}_S$ is lower than v_S. Hence we would have a contradiction to Lemma 1 if the outcome in $\mathcal{O}(\hat{v}_S, v_B)$ were no trade. □

We can now complete the proof of Proposition 2. Using Lemma 1 we have: $t^* = \underline{t}(\hat{v}_S, \hat{v}_B)$. By Claim 3, $t^* \leqslant \hat{v}_B$. Using Lemma 1 we also have: $t^* = \bar{t}(v_S, v_B)$. But then Lemma 1 and $t^* \leqslant \hat{v}_B$ implies that among the outcomes in $\mathcal{O}(\hat{v}_B, v_S)$ there must be a trade at a price below $\hat{v}_B$. This contradicts Claim 6. □