Carroll, Gabriel

Article

On mechanisms eliciting ordinal preferences

Theoretical Economics

Provided in Cooperation with:
The Econometric Society

Suggested Citation: Carroll, Gabriel (2018) : On mechanisms eliciting ordinal preferences, Theoretical Economics, ISSN 1555-7561, The Econometric Society, New Haven, CT, Vol. 13, Iss. 3, pp. 1275-1318, https://doi.org/10.3982/TE2774

This Version is available at:
https://hdl.handle.net/10419/197177

Theoretical Economics 13 (2018), 1275–1318 1555-7561/20181275

On mechanisms eliciting ordinal preferences

Gabriel Carroll
Department of Economics, Stanford University

When is a mechanism designer justified in only asking for ordinal information about preferences? Simple examples show that even if the planner's goal (expressed by a social choice correspondence (SCC)) depends only on ordinal information, eliciting cardinal information may help with incentives. However, if agents may be uncertain about their own cardinal preferences, then a strong robustness requirement can justify the focus on ordinal mechanisms. Specifically, when agents' preferences over pure outcomes are strict, if a planner is able to implement an SCC (in ex post equilibrium) using a mechanism that is robust to interdependence of arbitrary form in cardinal preferences, then there must exist such a mechanism that elicits only ordinal preferences. The strictness assumption can be dropped if we further allow the possibility of non-expected-utility preferences.

Keywords. Cardinal extension, ex post implementation, interdependence, ordinal mechanism, robust mechanism design.

JEL classification. D81, D82.

1. Introduction

In recent years, there has been growing interest in mechanism design problems in which agents report an ordinal preference ranking of outcomes to a central mechanism, and the mechanism chooses, as a function of these rankings, a lottery over outcomes. Some examples of such problems are as follows.

- Bogomolnaia and Moulin (2001) consider a problem of allocating n heterogeneous objects to n agents, one object per agent. Each agent reports a ranking over the objects, and the mechanism outputs a random allocation of objects. They prove an impossibility result: there is no strategy-proof mechanism satisfying certain efficiency and fairness criteria—specifically, ordinal efficiency and equal treatment of equals. A large subsequent literature has studied similar random matching problems, spurred partly by applications to school choice; see, e.g., Erdil (2014), Katta and Sethuraman (2006), Pycia and Ünver (2015).

Gabriel Carroll: gdc@stanford.edu

Thanks to (in random order) Juuso Toikka, Alex Wolitzky, Muhamet Yildiz, Yuichiro Kamada, Bengt Holmström, and Daron Acemoglu, and especially to Parag Pathak, for helpful discussions and advice. Suggestions from a co-editor and three referees also improved the paper. Dan Walton provided valuable research assistance.

 https://doi.org/10.3982/TE2774

- Random mechanisms for voting problems have been studied since Gibbard (1977), who considered agents reporting strict rankings over a finite set of outcomes and a mechanism choosing a lottery over the outcomes. Gibbard showed that any strategy-proof voting mechanism can be decomposed as a randomization over components that either are dictatorial or have a range of size two. Gibbard considered the unrestricted preference domain. Recent years have seen a surge of similar work on other preference domains, with both possibility and impossibility results. For example, Chatterji et al. (2014) give broad connectedness-type conditions on a domain that imply every strategy-proof and unanimous mechanism is a random dictatorship. They also show by example that domain conditions known to imply dictatorship results in the deterministic setting are not sufficient in the randomized setting. Other work in this vein includes Ehlers et al. (2002), Chatterji et al. (2016), Peters et al. (2014), Chatterji and Zeng (2018).
- Randomization has also been studied in rationing problems (Ehlers 2002, Ehlers and Klaus 2003).

Randomization is typically motivated by fairness concerns—for example, in school choice, lotteries can break ties to determine who gets a spot at a highly coveted school (Abdulkadiroğlu et al. 2005, Pathak and Sethuraman 2011)—although (as we shall see) it can also be helpful for incentives.

A key assumption throughout this literature is that agents can only report ordinal preference information (and the mechanism should provide incentives to do so truthfully). This assumption is usually justified by an informal appeal to simplicity (Bogomolnaia and Moulin 2001): it may be relatively easy for agents to express a preference ranking over pure outcomes, but harder to think about their own preferences over lotteries.

Still, we can easily think of mechanisms where agents are forced to make a choice between lotteries; ruling out such mechanisms is a substantive restriction. Indeed, another branch of school choice literature has emphasized the possible gains from mechanisms that elicit information about cardinal preferences (Abdulkadiroğlu et al. 2011, Abdulkadiroğlu et al. 2015, Pycia 2014, Troyan 2012). So we ask, "Is it possible to give a firm theoretical justification for only considering ordinal mechanisms?"

It is clear that if a social planner evaluates outcomes by a cardinal criterion (say, utilitarian welfare), then she should elicit cardinal preference information. In this paper, we therefore narrow down the question as follows: Suppose that the planner has a goal that only depends on agents' ordinal preferences. Might it nonetheless happen that, for incentive reasons, she can implement her goal only by having agents make choices that depend on cardinal preferences? Alternatively, is it indeed without loss of generality to assume an ordinal mechanism?

To set the stage, we give a simple example of how the restriction to ordinal mechanisms can matter, even if the planner's goals depend only on ordinal preferences.

EXAMPLE 1. Suppose there is just one agent, and four possible outcomes a, b, c, and d. Assume the agent evaluates lotteries by expected utility. The agent's ranking of pure

$t : a \succ d \succ b \succ c$	$\frac{1}{2}a + \frac{1}{2}d$
$t' : b \succ c \succ a \succ d$	$\frac{1}{2}b + \frac{1}{2}c$
$t'' : a \succ b \succ c \succ d$	$\frac{1}{2}a + \frac{1}{2}d$, $\frac{1}{2}b + \frac{1}{2}c$

TABLE 1. Preferences for Example 1

outcomes is known to fall under one of three possible ordinal types t, t', and t'', which rank the outcomes as shown in Table 1.

As shown, if the agent's ordinal type is t, we assume the planner wishes to implement the lottery over outcomes $\frac{1}{2}a + \frac{1}{2}d$. If the agent's type is t', the planner wishes to choose lottery $\frac{1}{2}b + \frac{1}{2}c$, and if the agent's type is t'', the planner is content with either of these two lotteries. (For brevity, we are just assuming the planner exogenously has these objectives; if we wished, we could "microfound" them in terms of fairness and efficiency criteria by introducing additional agents.)

Evidently, the planner's goal is a function only of the agent's ordinal preferences, and the goal can be achieved by simply asking the agent to choose between the two lotteries $\frac{1}{2}a + \frac{1}{2}d$ and $\frac{1}{2}b + \frac{1}{2}c$. But if the planner is restricted to use a direct mechanism that asks the agent to report his ordinal preferences, she is out of luck: Either the mechanism must specify that type t'' gets lottery $\frac{1}{2}a + \frac{1}{2}d$ or that he gets $\frac{1}{2}b + \frac{1}{2}c$. In each case, the mechanism fails to provide incentives to report truthfully; there will be at least some cardinal preferences consistent with type t'' for which the agent will prefer to misreport his type as t or t'. ◊

This example shows that, without further assumptions, it is not without loss of generality to restrict to ordinal mechanisms.

The main contribution of the present paper is a formal framework in which the restriction *is* without loss of generality. As in the example, our planner who has some goals that depend on agents' ordinal preferences (represented by a *social choice correspondence* (SCC), specifying the acceptable outcome lotteries for each preference profile). She wishes to know whether it is possible to implement her goals. This setting embeds concrete applications of interest. We return to our opening examples.

- In the object allocation problem of Bogomolnaia and Moulin (2001), the requirements of ordinal efficiency and equal treatment of equals can be expressed by an SCC that, at each profile of ordinal preferences, deems a lottery over allocations to be acceptable if and only if it is consistent with these requirements.

- In a voting problem, unanimity can be expressed by the SCC where, at any profile where all voters have the same preferred outcome, only the degenerate lottery on that outcome is acceptable, and at any other preferences, all lotteries are acceptable.

  Since our framework is specifically focused on possibility/impossibility results, it does not capture characterizations such as the random dictatorship result of Chatterji et al. (2014). However, one could refine the SCC so as to rule out random dictatorship (for example, by imposing a "compromise" requirement as in

Chatterji et al. (2016), specifying that at some particular profiles, an outcome that is not anyone's top choice should be chosen with some minimum probability). Then the result of Chatterji et al. (2014) translates into an impossibility theorem, which fits within our framework.

Our planner can consider her problem in two ways. She can ask whether her goals are implementable by some mechanism. She can also ask specifically whether they can be implemented by an ordinal mechanism. When the answers to these two questions are equivalent, the planner is justified in restricting her attention to ordinal mechanisms. We then say that there is a *foundation for ordinal mechanisms*.

Our main result, Theorem 1, gives such a foundation, showing that the planner can implement her goals in a way that is sufficiently robust to uncertainty about cardinal preferences only if she can implement them using an ordinal mechanism. At the same time, a series of examples illustrates the limits of the argument.

A crucial ingredient in this foundation is *interdependence* in cardinal preferences: while each agent knows his own ordinal preferences, we allow that his preferences over lotteries might depend on other agents' private information. While the introduction of interdependence may appear surprising, there are two arguments for why it is reasonable, as we elaborate in Section 4. First, it arises naturally if we attempt to model agents not perfectly knowing their own cardinal preferences: this uncertainty may be correlated with others' preferences, generating interdependence. Second, it is realistic to suppose that some modest amount of interdependence exists in applications. Our foundation in fact applies even if the amount of interdependence is restricted to be arbitrarily small, as long as it is nonzero.

We use ex post implementation (Bergemann and Morris 2005, Chung and Ely 2006, Jehiel et al. 2006) as our solution concept. This is motivated by the existing literature on probabilistic mechanism design with ordinal preferences (such as that cited above), which typically requires mechanisms to be strategy-proof; i.e., reporting truthfully should be a dominant strategy for each agent i and all von Neumann–Morgenstern utility functions consistent with i's ordinal preferences. The restriction to dominant-strategy mechanisms has been critiqued elsewhere, in an influential paper by Bergemann and Morris (2005) (which we discuss further momentarily), but it is not our focus here. So we simply follow the relevant literature and retain this solution concept; adapting it for interdependent preferences leads us to ex post implementation.[1]

A limitation of our Theorem 1 is that it only applies when ordinal preferences over pure outcomes are strict. This means that while the theorem applies to the voting example above, it does not apply (for example) to matching settings, where each agent is indifferent among all allocations that gave him the same object.

This limitation can be overcome if we expand our model further. Specifically, we note that ordinal mechanisms are not only robust to arbitrary cardinal preferences, but remain robust even if agents' preferences over lotteries are not represented by expected utility at all. This leads us to consider a planner who wants to accommodate agents

[1] Section 6.2 contains some discussion of Bayesian implementation.

with non-expected-utility preferences. We can then give a foundation for ordinal mechanisms without assuming strict preferences, either when preferences over lotteries are unrestricted (Theorem 2) or when they fall within any of several commonly studied non-expected-utility models (Theorem 3). Interdependence is still needed.

A quick overview of the paper is as follows. Example 1 above illustrates the basic challenge in looking for foundations for ordinal mechanisms. After we set up the general framework in Section 2, we first consider one possible response to the challenge: to try to give a foundation for ordinal mechanisms by restricting the SCCs under consideration. Section 3 proposes one natural restriction, and offers examples to show how the restriction does not help in general. This motivates the alternative response, of strengthening the robustness requirement by allowing interdependence in cardinal preferences. In Section 4, we give the main result (Theorem 1), the foundation for ordinal mechanisms with interdependence when preferences are strict. In Section 5, we proceed to allow non-expected utility, and give the foundations for ordinal mechanisms without strict preferences (Theorems 2 and 3). After briefly discussing extensions, in the concluding section, we discuss some more general issues.

This paper owes a methodological debt to the work of Bergemann and Morris (2005). They studied the question of whether implementability of an SCC using a mechanism in which agents' beliefs play no role (i.e., ex post implementation) is equivalent to implementability over all possible belief hierarchies. Analogously, we ask here whether implementability using a mechanism in which cardinal preferences play no role is equivalent to implementability for all possible cardinal preferences.

Our Theorem 1 also is reminiscent of the result of Jehiel et al. (2006). They consider ex post implementation in a setting with interdependent preferences and monetary transfers with quasilinear utility. Like this paper, theirs shows that implementation is typically not possible unless it is possible for a trivial reason: in their paper, the trivial case is a constant mechanism. Their paper and this one differ both in the setting and in the breadth of the class of SCCs considered. It does not seem that our result follows from that of Jehiel et al. (2006) or vice versa.

One other closely related paper is by Ehlers et al. (2016). That paper, independent of the present work, also seeks justification for ordinal mechanisms. It considers mechanisms defined over all possible cardinal preferences consistent with a given domain of (strict) ordinal preferences, and shows that if an incentive-compatible mechanism satisfies a uniform continuity condition within each ordinal type, it must actually be an ordinal mechanism. The uniform continuity hypothesis imposed there is a strong requirement: indeed, if it were required across *all* cardinal preferences of a given agent, rather than just cardinal preferences with the same ordinal preference, the mechanism would have to be constant.

Finally, this paper ties in with the broader topic of deciding what message spaces are appropriate when designing a mechanism—an issue that has recently gained prominence in market design practice (Milgrom 2011). The paper is written with a focus on ordinal versus cardinal preference information, but, in fact, the same methods can be applied more generally to other kinds of preference information to ask when such information can be safely excluded from the message space by a mechanism designer who does not directly care about it. (This is discussed further in Section 6.1.)

2. The modeling framework

We now introduce the necessary definitions to lay out the formal framework and state the main result, Theorem 1. Definitions needed for other results and extensions, including the material on non-expected utility in Section 5, are introduced later as needed.

We assume a set $N = \{1, \dots, n\}$ of *agents* and a finite set X of *outcomes* are given. In a voting context, outcomes might correspond to candidates or policies; in a matching context, each element of X would be a possible matching. We write $\Delta(X)$ for the space of lotteries over X, and write $\pi(x)$ for the probability of outcome x under lottery π.

An *ordinal type space* for agent i is a set T_i, where each $t_i \in T_i$ is a weak ordering (i.e., a complete, transitive ordering) over X. An element $t_i \in T_i$ is called an *ordinal type*. We write $x \succeq_{t_i} y$ to indicate that t_i ranks x weakly above y, and write $x \succ_{t_i} y$ and $x \sim_{t_i} y$ similarly. We also use the term *ordinal type space* for the product $T = T_1 \times \cdots \times T_n$. The definition implicitly imposes that different types must have different preference orderings; this is not a substantive restriction. We write $T_{-i} = \times_{j \neq i} T_j$. We say that *types are strict* if, for all t_i and all distinct $x, y \in X$, $x \nsim_{t_i} y$.

Next, we need to describe the goals of the social planner. These goals are represented by a *social choice correspondence* (SCC). We take as given that the domain of the SCC is the space of ordinal types T. Thus, an SCC is a correspondence $F : T \rightrightarrows \Delta(X)$, specifying an acceptable set of lotteries for each type profile $t \in T$. We assume $F(t) \neq \varnothing$ for all t.

We say that F is *deterministic* if, for each t, $F(t) \subseteq X$: that is, F only allows degenerate lotteries. We say that F is a *social choice function* (SCF) if it is single-valued.

Given a planner's goals, as represented by the SCC F, there are two natural ways to ask whether F can be implemented:

(Q1) Can F be implemented by some ordinal mechanism?

(Q2) Can F always be implemented by some mechanism, in spite of any uncertainty about agents' preferences over lotteries?

We discuss in a moment how to state these questions formally. But first we comment that in any reasonable modeling framework, an affirmative answer to (Q1) should imply an affirmative answer to (Q2), since the class of mechanisms allowed by (Q1) is a subset of those in (Q2). The general question we wish to answer is the converse: whether an affirmative answer to (Q2) implies an affirmative answer to (Q1). If we can formalize (Q1) and (Q2) in such a way that this converse holds, we will say that we have a *foundation for ordinal mechanisms*.

To formalize (Q1), we just need to define ordinal mechanisms. Given a ordinal type space T, an *ordinal mechanism* is a function $M : T \to \Delta(X)$. We say that M *implements* F *(in dominant strategies)* if the following criteria hold:

- For each type profile t, $M(t) \in F(t)$.
- For all i and all $t_i, t_i' \in T_i$ and $t_{-i} \in T_{-i}$, the lottery $M(t_i, t_{-i})$ first-order stochastically dominates $M(t_i', t_{-i})$ with respect to the preference ordering t_i; that is, for all

$x \in X$,

$$\sum_{y \succeq_{t_i} x} M(t_i, t_{-i})(y) \geq \sum_{y \succeq_{t_i} x} M\big(t_i', t_{-i}\big)(y).$$

One readily checks that the latter condition is equivalent to requiring that every von Neumann–Morgenstern utility function representing t_i weakly prefers $M(t_i, t_{-i})$ to $M(t_i', t_{-i})$, and so is the appropriate formulation of dominant-strategy incentive compatibility when agents report ordinal preferences.

We say that the mechanism M is *deterministic* if $M(t)$ is a degenerate lottery for each t. Note that a mechanism M implementing F may be deterministic even if F itself is not deterministic.

To formalize (Q2), the first natural framework to use is one of private values; that is, agents know their own cardinal (von Neumann–Morgenstern) utilities. We then define the relevant spaces of uncertain preferences as follows: An *(independent) cardinal type space* for agent i is a finite set S_i of possible von Neumann–Morgenstern utility functions $s_i : X \to \mathbb{R}$. (Finiteness is again not a substantive restriction.) A type $s_i \in S_i$ *represents* an ordinal type t_i if, for all $x, y \in X$, we have $s_i(x) \geq s_i(y)$ if and only if $x \succeq_{t_i} y$. For a lottery π, we write $s_i(\pi)$ for the expected utility $\sum_{x \in X} \pi(x) s_i(x)$. We say that S_i is a *cardinal extension* of T_i if every $s_i \in S_i$ represents some type in T_i, and every type in T_i is represented by some type in S_i.

We say that the cardinal type space $S = S_1 \times \cdots \times S_n$ is a *cardinal extension* of T if S_i is a cardinal extension of T_i for each i. In this case, we write $t_i(s_i)$ for the ordinal type represented by s_i, and write $t(s)$ for the profile of ordinal types $(t_1(s_1), \ldots, t_n(s_n))$. The overloaded notation t_i (representing both an ordinal type and a function $S_i \to T_i$) should not cause any confusion in practice. We also again write $S_{-i} = \times_{j \neq i} S_j$.

We also say that S_i is a *minimal* cardinal extension of T_i if each $t_i \in T_i$ is represented by just one $s_i \in S_i$. Minimal cardinal extensions are the formal analogue in our setting of the *payoff type spaces* considered by Bergemann and Morris (2005).

We define a *cardinal mechanism* over S to be a function $M : S \to \Delta(X)$. If S is a cardinal extension of T, we can say that M *implements* F *(in dominant strategies)* if the following statements hold:

- For each s, $M(s) \in F(t(s))$.
- For all i, $s_i, s_i' \in S_i$ and all $s_{-i} \in S_{-i}$,

$$s_i\big(M(s_i, s_{-i})\big) \geq s_i\big(M\big(s_i', s_{-i}\big)\big).$$

Cardinal mechanisms give more flexibility than ordinal mechanisms, by allowing the chosen lottery $M(s)$ to vary depending on agents' cardinal preferences, even within a fixed profile of ordinal preferences.

The relevant formulation of (Q2) in this framework is then, "Can F be implemented over every cardinal extension of the given type space T?"

In Example 1, the answer to (Q1) was no, while the answer to (Q2) (as formulated above) was yes.

The above framework shows one way to formalize (Q2). But our main result requires larger spaces of uncertain preferences: it requires a framework in which each agent i is uncertain about his own cardinal preferences, and knowing other agents' types would be informative in resolving this uncertainty. That is, it requires interdependence in cardinal utilities. We give the definitions here, leaving the detailed discussion of interpretation to Section 4.

Define an *interdependent cardinal type space* to be a pair (S, u), whose components are as follows:

- The set $S = S_1 \times \cdots \times S_n$ is a product of (finite) sets representing the possible types of the individual agents.
- The object u is a profile of utility functions, $u = (u_1, \ldots, u_n)$, where each function $u_i : X \times S \to R$ expresses agent i's utility for each outcome in X, which may depend on *all* agents' types.

We again extend utilities to lotteries linearly, writing $u_i(\pi, s) = \sum_{x \in X} \pi(x) u_i(x, s)$, the expected utility of lottery π when the type profile is s.

We say that (S, u) is an *interdependent cardinal extension* of the ordinal type space T if there are surjective functions $t_i : S_i \to T_i$ for each i, such that for all type profiles $s \in S$ and all agents i, $u_i(\cdot, s)$ represents the ordinal type $t_i(s_i)$. The extension is *minimal* if each function t_i is a bijection.

The appropriate analogue of dominant-strategy implementation when preferences are interdependent is ex post implementation (Chung and Ely 2006), which requires only that it should be optimal for i to report his type truthfully regardless of the true type profile, as long as other agents are also reporting their types truthfully. Accordingly, when (S, u) is an interdependent cardinal extension of T, we define a *mechanism* over (S, u) to be a function $M : S \to \Delta(X)$, and we say that M *implements* F *(in ex post equilibrium)* if the following statements hold:

- For each s, $M(s) \in F(t(s))$.
- For all i, $s_i, s_i' \in S_i$ and all $s_{-i} \in S_{-i}$,

$$u_i\big(M(s_i, s_{-i}), s_i, s_{-i}\big) \geq u_i\big(M\big(s_i', s_{-i}\big), s_i, s_{-i}\big).$$

The statement of (Q2) in this interdependent framework is then, "Can F be implemented over every interdependent cardinal extension of T?"

Of course, if the answer to (Q2) in the interdependent framework is affirmative, then the answer in the (less demanding) independent framework is also affirmative.

We close the section with a simple observation that helps narrow the focus of our study.

PROPOSITION 1. *Suppose that either of the following statements holds:*

(a) The SCC F is a social choice function.

(b) The SCC F is deterministic.

If F can be implemented over every (independent) cardinal extension of T, then it can be implemented by an ordinal mechanism.[2]

The (straightforward) proof is provided in Appendix A. (Part (a) parallels a similar observation by Bergemann and Morris 2005, their Proposition 2.)

Thus, in these two cases, we have a ready foundation for ordinal mechanisms. So the challenge in giving foundations pertains specifically to situations where the SCC F is not single-valued and genuinely involves lotteries.

One more comment about the modeling: since the goal is to give microfoundations by having agents' preferences fully modeled in the type space, why keep dominant-strategy (respectively, ex post) implementation as the solution concept, instead of modeling agents' beliefs about each other as part of the type and using Bayesian implementation? Our main answer, again, is that this critique of ex post implementation has been raised elsewhere—by Bergemann and Morris (2005) as well as others (Chung and Ely 2007, Yamashita 2015)—and is not the focus of this paper. We keep ex post so as to concentrate attention on the restriction to ordinal mechanisms. In addition, there does not seem to be a sensible way to incorporate interdependence if we adopt Bayesian implementation; see Section 6.2.

3. Restricted SCC's and examples

We already saw, in Example 1, that there is no fully general foundation for ordinal mechanisms. A desire for robustness to uncertain cardinal preferences does not immediately justify the use of ordinal mechanisms.

This suggests two possible routes to look for such foundations. The first route is to impose restrictions on the structure of the SCC F. The second is to reformulate (Q2) by requiring robustness to larger spaces of uncertainty.

The rest of this section considers one version of the first route. The approach we take here is not used for our main results later, so it can be skipped on a quick reading; however, this section also develops examples that are referred to again in later exposition. The following section undertakes the second route, by allowing for interdependent preferences.

To rule out the problem encountered in Example 1, we consider restricting the SCC F as follows: for each t, require that $F(t) = \Delta(Y)$ for some $Y \subseteq X$. Such an F is called *simple*. This represents a situation in which, for each ordinal type profile, every pure outcome is considered either acceptable or unacceptable, and the planner simply wishes to be sure of an acceptable outcome. Criteria such as unanimity in voting problems or ex post Pareto efficiency in matching problems are expressed by a simple SCC.

Under the requirement of simplicity, there is, in fact, a foundation for ordinal mechanisms with just one agent.

Proposition 2. *Suppose $n = 1$. If the simple SCC F is implementable over every cardinal extension of T_1, then it is implementable by an ordinal mechanism over T_1.*

[2] Of course, the statement remains true if we allow interdependence, since implementability over every interdependent extension is a stronger hypothesis.

The proof involves giving a simple iterative-elimination algorithm that determines whether F is implementable by an ordinal mechanism over T_1. If not, we can use the execution path of the algorithm to explicitly construct a cardinal extension over which F is not implementable. The details of the proof are given in Appendix A. (In fact, the proof shows a stronger statement: either F is implementable by a *deterministic* ordinal mechanism or there is a *minimal* cardinal extension over which it is not implementable.)

Unfortunately, Proposition 2 does not extend beyond the one-agent case.

The intuitive reason is that the simplicity restriction on F loses its bite when there are multiple agents: The interaction between different agents' incentive constraints can force the use of lotteries, even though the definition of F itself does not require it. This fact is illustrated by Example 2 below. Afterward, we show how this idea can be developed into a counterexample to Proposition 2 with multiple agents.

EXAMPLE 2. Let X consist of eight outcomes $\{a, b, c, d, e, f, g, h\}$, and let there be two agents, each with two ordinal types. Let the preferences of the types be

$$t_1 : b \succ f \succ a \succ e \succ c \succ g \succ d \succ h,$$
$$t_1' : h \succ d \succ g \succ c \succ e \succ b \succ f \succ a,$$
$$t_2 : a \succ c \succ b \succ d \succ f \succ h \succ e \succ g,$$
$$t_2' : g \succ e \succ h \succ f \succ d \succ a \succ c \succ b.$$

Suppose the simple SCC F specifies two acceptable outcomes for each type profile:

	t_2	t_2'
t_1	a, b	c, d
t_1'	e, f	g, h

It is straightforward to check that F is implemented by the ordinal mechanism specifying $\frac{1}{2} - \frac{1}{2}$ lotteries for each type profile, i.e.,

	t_2	t_2'
t_1	$\frac{1}{2}a + \frac{1}{2}b$	$\frac{1}{2}c + \frac{1}{2}d$
t_1'	$\frac{1}{2}e + \frac{1}{2}f$	$\frac{1}{2}g + \frac{1}{2}h$

We claim that this is the *only* ordinal mechanism implementing F. Indeed, let M be such a mechanism. Consider the total probability of obtaining outcomes weakly preferred to f by type t_1 (namely b, f) at each of the profiles (t_1, t_2) and (t_1', t_2). The incentive constraint of type t_1 gives

$$M(t_1, t_2)(b) + M(t_1, t_2)(f) \geq M\big(t_1', t_2\big)(b) + M\big(t_1', t_2\big)(f).$$

Since F prohibits M from placing positive probability on f at (t_1, t_2) or on b at (t_1', t_2), we obtain more simply $M(t_1, t_2)(b) \geq M(t_1', t_2)(f)$. We summarize the above reasoning as

$$(t_1, t_2) \to \big(t_1', t_2\big), \text{outcomes} \succeq f : \qquad M(t_1, t_2)(b) \geq M\big(t_1', t_2\big)(f).$$

Using the other incentive constraints for each agent, we similarly obtain the inequalities

$$\begin{aligned}
(t_1', t_2) \to (t_1', t_2'), \text{outcomes} \succeq h: &\qquad M(t_1', t_2)(f) \geq M(t_1', t_2')(h),\\
(t_1', t_2') \to (t_1, t_2'), \text{outcomes} \succeq d: &\qquad M(t_1', t_2')(h) \geq M(t_1, t_2')(d),\\
(t_1, t_2') \to (t_1, t_2), \text{outcomes} \succeq a: &\qquad M(t_1, t_2')(d) \geq M(t_1, t_2)(a),\\
(t_1, t_2) \to (t_1, t_2'), \text{outcomes} \succeq c: &\qquad M(t_1, t_2)(a) \geq M(t_1, t_2')(c),\\
(t_1, t_2') \to (t_1', t_2'), \text{outcomes} \succeq g: &\qquad M(t_1, t_2')(c) \geq M(t_1', t_2')(g),\\
(t_1', t_2') \to (t_1', t_2), \text{outcomes} \succeq e: &\qquad M(t_1', t_2')(g) \geq M(t_1', t_2)(e),\\
(t_1', t_2) \to (t_1, t_2), \text{outcomes} \succeq b: &\qquad M(t_1', t_2)(e) \geq M(t_1, t_2)(b).
\end{aligned}$$

This cycle of eight inequalities immediately means that all the probabilities involved must be equal; thus, they are all equal to $1/2$, and the mechanism M is uniquely determined, as claimed. ◊

Now, we are ready to see how, even with the restriction to simple SCCs, we do not have a foundation for ordinal mechanisms in general: Example 3 gives a simple SCC that can be implemented over every cardinal extension, but not implemented by an ordinal mechanism. It draws on the construction from Example 2 to ensure that any ordinal mechanism would need to use lotteries; once this door is opened, we can then use the idea of Example 1 to force one agent to choose among lotteries in a way that depends on cardinal preferences.

Example 3. Let X consist of 11 outcomes $\{a, a', b, b', c, d, e, f, g, h, m\}$. Let there be three agents; agents 1 and 2 have two possible ordinal types each, while agent 3 has three possible ordinal types. The preferences of the types are

$$\begin{aligned}
&t_1 : b \succ b' \succ f \succ a \succ a' \succ e \succ c \succ g \succ d \succ h \succ m,\\
&t_1' : m \succ h \succ d \succ g \succ c \succ e \succ b \succ b' \succ f \succ a \succ a',\\
&t_2 : a \succ a' \succ c \succ b \succ b' \succ d \succ f \succ h \succ e \succ g \succ m,\\
&t_2' : m \succ g \succ e \succ h \succ f \succ d \succ a \succ a' \succ c \succ b \succ b',\\
&t_3 : a \succ b' \succ b \succ a' \succ c, d, e, f, g, h \succ m,\\
&t_3' : a' \succ a \succ b' \succ b \succ c, d, e, f, g, h \succ m,\\
&t_3'' : m \succ a \succ a' \succ b' \succ b \succ c, d, e, f, g, h.
\end{aligned}$$

(Here the commas in the preferences of agent 3 mean that we may choose preferences among $c, \ldots, h$ arbitrarily.)

Let F be the simple SCC whose acceptable outcomes at each type profile are

t_3:

	t_2	t_2'
t_1	a, b	c, d
t_1'	e, f	g, h

t_3':

	t_2	t_2'
t_1	a', b'	c, d
t_1'	e, f	g, h

t_3'':

	t_2	t_2'
t_1	a, b, a', b'	m
t_1'	m	m

First we check that F cannot be implemented by any ordinal mechanism. Suppose such a mechanism M exists. When agent 3 is of type t_3, the preferences of agents 1 and 2 over the feasible outcomes exactly replicate Example 2, and so M must prescribe the lotteries

t_3:

	t_2	t_2'
t_1	$\frac{1}{2}a + \frac{1}{2}b$	$\frac{1}{2}c + \frac{1}{2}d$
t_1'	$\frac{1}{2}e + \frac{1}{2}f$	$\frac{1}{2}g + \frac{1}{2}h$

Similarly, when agent 3 has type t_3', M must prescribe

t_3':

	t_2	t_2'
t_1	$\frac{1}{2}a' + \frac{1}{2}b'$	$\frac{1}{2}c + \frac{1}{2}d$
t_1'	$\frac{1}{2}e + \frac{1}{2}f$	$\frac{1}{2}g + \frac{1}{2}h$

Now consider the possible values for the lottery $M(t_1, t_2, t_3'')$. Fix the types of the first two agents at (t_1, t_2). We must have $M(t_1, t_2, t_3'')(a) \geq \frac{1}{2}$ or else type t_3'' would have a potential incentive to imitate t_3. However, $M(t_1, t_2, t_3'')(a) + M(t_1, t_2, t_3'')(a') \leq \frac{1}{2}$ or else the incentive constraint of type t_3' would be violated, and $M(t_1, t_2, t_3'')(a) + M(t_1, t_2, t_3'')(b') \leq \frac{1}{2}$ or else the incentive constraint of type t_3 would be violated. These inequalities imply that $M(t_1, t_2, t_3'')$ puts probability $\frac{1}{2}$ on a, and probability 0 on a' and b'. The remaining probability $\frac{1}{2}$ must go to b. But this gives t_3'' an incentive to imitate t_3', and we reach a contradiction.

Thus our F cannot be implemented by an ordinal mechanism. Nonetheless, it can be implemented over any cardinal extension of the type space T as follows. If agent 3's preferences correspond to ordinal type t_3 or t_3', then execute the $\frac{1}{2} - \frac{1}{2}$ lottery given by the agents' ordinal preference types, as above. Otherwise, the outcome is given by

t_3'':

	t_2	t_2'
t_1	$\frac{1}{2}a + \frac{1}{2}b$ or $\frac{1}{2}a' + \frac{1}{2}b'$	m
t_1'	m	m

Here the interpretation of the "or" is that if the ordinal types are (t_1, t_2, t_3''), the mechanism prescribes whichever of the two indicated lotteries is preferred by agent 3—a choice that depends on 3's cardinal preferences. ◊

4. Foundations with interdependent preferences

We now take the second route described at the beginning of Section 3: we enlarge the space of uncertainty by allowing agents' cardinal preferences to be interdependent. In this framework, we can give a foundation for ordinal mechanisms. This is our main result. We also present some examples that show the role of various hypotheses. Modeling issues are discussed afterward in Section 4.2.

4.1 *Formal results*

When there is only one agent, interdependence buys us nothing; hence, we had better assume $n \geq 2$ agents. In this case, we can indeed obtain a foundation for ordinal mechanisms. The only assumption we need to make about F is that it is closed-valued. However, we do need to assume that the agents' ordinal preferences over pure outcomes are strict.

Theorem 1. *Suppose $n \geq 2$ and that T is an ordinal type space in which all agents' types are strict. Let $F : T \rightrightarrows \Delta(X)$ be an SCC with $F(t)$ closed for each $t \in T$. If F is implementable over every interdependent cardinal extension of T, then it is implementable by an ordinal mechanism over T.*

We sketch the proof here by presenting the argument in the case of two agents and three outcomes. The full proof uses similar ideas, but is more notationally involved and is left to Appendix A.[3]

Proof of Theorem 1: Sketch (Two agents, three outcomes). For any $\epsilon > 0$, say that the ordinal mechanism $M : T \to \Delta(X)$ *ϵ-implements* F over T if the following statements hold:

- For each $t \in T$, $M(t) \in F(t)$. and
- For each i, all $t_i, t_i' \in T_i$ and $t_{-i} \in T_{-i}$, and all outcomes $x \in X$,

$$\sum_{y \succeq_{t_i} x} M(t_i, t_{-i})(y) \geq \sum_{y \succeq_{t_i} x} M\big(t_i', t_{-i}\big)(y) - \epsilon. \tag{1}$$

Fix ϵ. We construct an interdependent cardinal extension S of T, such that if F is implementable over S, then F is ϵ-implementable by an ordinal mechanism. Exact implementability then follows by a limiting argument taking $\epsilon \to 0$ (and using the fact that $F(t)$ is closed).

As a matter of notation, we identify utility functions with elements of the vector space $\mathbb{R}^3$; note that expected utility for a lottery $u(\pi)$ is then given by the inner product $u \cdot \pi$.

[3]The sketch here is not quite a specific instance of the general proof in the Appendix; it employs some extra shortcuts to make the argument more compact in this special case.

	s_2	s_2'
s_1	u_1 u_2	$u_1' - w$ $u_2 + w$
s_1'		u_1' u_2'

FIGURE 1. Basic building block for the interdependent type space.

The basic building block of the construction is shown in Figure 1. Here s_1 and s_1' are two cardinal types that represent the same ordinal type t_1 of agent 1. We take $u_1, u_1' : X \to \mathbb{R}$ to be any two different utility functions that represent this type, and similarly for agent 2. We also consider an arbitrary $w \in \mathbb{R}^3$, close enough to 0 so that $u_1' - w$ still represents t_1 and $u_2 + w$ still represents t_2. (This holds whenever w is small enough, since types are strict.) Then at the three interdependent cardinal type profiles (s_1, s_2), (s_1, s_2'), and (s_1', s_2'), we specify cardinal preferences over X for the two agents as indicated. At (s_1', s_2), the cardinal preferences may be arbitrary.

Suppose a mechanism M implements F over an interdependent cardinal extension that contains this building block. There are two ex post incentive constraints for agent 2 when agent 1's type is s_1: one for s_2 to imitate s_2' and one for s_2' to imitate s_2. These give

$$u_2 \cdot M(s_1, s_2) \geq u_2 \cdot M\big(s_1, s_2'\big), \qquad (u_2 + w) \cdot M\big(s_1, s_2'\big) \geq (u_2 + w) \cdot M(s_1, s_2).$$

Subtracting these inequalities gives $w \cdot (M(s_1, s_2') - M(s_1, s_2)) \geq 0$. A similar calculation using agent 1's incentive constraints gives $w \cdot (M(s_1', s_2') - M(s_1, s_2')) \geq 0$. Combining gives $w \cdot (M(s_1', s_2') - M(s_1, s_2)) \geq 0$.

The next step of the construction is to combine several of these blocks into the staircase shape in Figure 2. For each agent $i = 1, 2$, we now start from seven arbitrary utility functions $u_i^1, \ldots, u_i^7$ that all represent the same t_i. We also take three vectors

$$w^1 = (\delta, -\delta, 0), \qquad w^2 = (0, \delta, -\delta), \qquad w^3 = (-\delta, 0, \delta),$$

where, as before, $\delta > 0$ is small enough so that each $u_i^j \pm w^d$ represents t_i. To form the staircase, we define seven cardinal types $s_i^1, \ldots, s_i^7$ for each agent i and specify cardinal preferences as indicated in the figure. Again, for the cells left blank, we can specify any cardinal preferences as long as they are consistent with the ordinal types t_1 and t_2.

Suppose that a mechanism M implements F over an extension in which this staircase appears. Notice that for each $k = 1, \ldots, 6$, the types s_1^k and s_1^{k+1} for agent 1, and s_2^k and s_2^{k+1} for agent 2 form a building block as in Figure 1. Using this block, we conclude that $w^1 \cdot (M(s_1^{k+1}, s_2^{k+1}) - M(s_1^k, s_2^k)) \geq 0$. Adding together these inequalities for each $k = 1, \ldots, 6$ gives

$$w^1 \cdot \big(M\big(s_1^7, s_2^7\big) - M\big(s_1^1, s_2^1\big)\big) \geq 0. \tag{2}$$

Now, for each k, the types s_1^k, s_1^{k+2} and s_2^k, s_2^{k+2} also form a building block, with perturbation w^2 rather than w^1. So $w^2 \cdot (M(s_1^{k+2}, s_2^{k+2}) - M(s_1^k, s_2^k)) \geq 0$. Adding up for

	s_2^1	s_2^2	s_2^3	s_2^4	s_2^5	s_2^6	s_2^7
s_1^1	u_1^1 u_2^1	$u_1^2 - w^1$ $u_2^1 + w^1$	$u_1^3 - w^2$ $u_2^1 + w^2$	$u_1^4 - w^3$ $u_2^1 + w^3$			
s_1^2		u_1^2 u_2^2	$u_1^3 - w^1$ $u_2^2 + w^1$	$u_1^4 - w^2$ $u_2^2 + w^2$	$u_1^5 - w^3$ $u_2^2 + w^3$		
s_1^3			u_1^3 u_2^3	$u_1^4 - w^1$ $u_2^3 + w^1$	$u_1^5 - w^2$ $u_2^3 + w^2$	$u_1^6 - w^3$ $u_2^3 + w^3$	
s_1^4				u_1^4 u_2^4	$u_1^5 - w^1$ $u_2^4 + w^1$	$u_1^6 - w^2$ $u_2^4 + w^2$	$u_2^7 - w^3$ $u_2^4 + w^3$
s_1^5					u_1^5 u_2^5	$u_1^6 - w^1$ $u_2^5 + w^1$	$u_1^7 - w^2$ $u_2^5 + w^2$
s_1^6						u_1^6 u_2^6	$u_1^7 - w^1$ $u_2^6 + w^1$
s_1^7							u_1^7 u_2^7

FIGURE 2. Staircase assembly of building blocks.

$k = 1, 3, 5$ gives

$$w^2 \cdot \big(M(s_1^7, s_2^7) - M(s_1^1, s_2^1)\big) \geq 0. \tag{3}$$

Similarly, types s_1^k, s_1^{k+3} and s_2^k, s_2^{k+3} form a building block with perturbation w^3, from which (taking $k = 1, 4$ and adding)

$$w^3 \cdot \big(M(s_1^7, s_2^7) - M(s_1^1, s_2^1)\big) \geq 0. \tag{4}$$

Adding (2), (3), and (4) gives $0 \geq 0$. So all three of them must be equalities: the vector $M(s_1^7, s_2^7) - M(s_1^1, s_2^1)$ is orthogonal to w^1, w^2, and w^3. It is also orthogonal to $(1, 1, 1)$, since it is the difference of two lotteries. Thus, $M(s_1^7, s_2^7) - M(s_1^1, s_2^1)$ is orthogonal to all of $\mathbb{R}^3$, so it must be zero. We conclude that

$$M(s_1^7, s_2^7) = M(s_1^1, s_2^1). \tag{5}$$

Now we are ready to construct a full-fledged interdependent cardinal extension S. For each agent i and each $t_i \in T_i$, define seven corresponding cardinal types $s_i^1(t_i), \ldots, s_i^7(t_i)$. Each ordinal type profile in T thus gives rise to 49 cardinal type profiles in S, and we fill in the cardinal preferences as in Figure 2. Since the cardinal utilities u_i^j there were arbitrary, we can add the following stipulations for each t_i:

- The function $u_i^1(t_i)$ should assign t_i's top-, middle-, and bottom-ranked outcomes the cardinal values 1, ϵ, and 0, respectively.

- The function $u_i^7(t_i)$ should assign t_i's top-, middle-, and bottom-ranked outcomes the cardinal values 1, $1-\epsilon$, and 0.

This specifies S. (Again, blank cells can be filled in freely.)

Now, by hypothesis, F can be implemented over S by a mechanism M. In view of the above analysis, and specifically (5), we can define an ordinal mechanism $M' : T \to \Delta(X)$ by

$$M'(t_1, t_2) = M\big(s_1^7(t_1), s_2^7(t_2)\big) = M\big(s_1^1(t_1), s_2^1(t_2)\big).$$

Our goal is to show that M' ϵ-implements F over T. But the ordinal incentive constraints (1) for M' just follow from the cardinal incentive constraints for M: Indeed, consider any $(t_1, t_2) \in T$ and any possible deviation t_1' for agent 1. The cardinal incentive constraint at type profile $(s_1^1(t_1), s_2^1(t_2))$, which specifies that agent 1 does not want to deviate to $s_1^1(t_1')$ in M, implies that in M', the probability of getting the top-ranked outcome cannot increase by more than ϵ if 1 misreports t_1'. Similarly, the cardinal incentive constraint at type profile $(s_1^7(t_1), s_2^7(t_2))$, guarding against misreport $s_1^7(t_1')$ in M, implies that in M', the probability of getting one of the two top outcomes cannot increase by more than ϵ under the misreport t_1'. Together, these constraints constitute the ordinal incentive constraints (1) for agent 1 in M', and similarly for agent 2.

Thus, M' ϵ-implements F over T. By a limiting argument (detailed in the full proof), we find that F can be exactly implemented over T. □

The key to the argument is the staircase construction that forces M to be equal across cardinal type profiles with different cardinal utility functions, as indicated by (5). This makes it possible to replicate all of the (ϵ-) ordinal incentive constraints within the cardinal type space. Of course, this construction leans heavily on the fact that we have allowed interdependence. Without interdependence, we could not construct the staircase as in Figure 2, since agent 1's utility function over X would have to be the same across all cells in any given row, and 2's utility would have to be the same across all cells in any given column.

We next give a couple of examples to show that various hypotheses in Theorem 1 cannot be dropped.

Example 4 (Hypothesis of strict preferences). If we allow for weak preferences, then Theorem 1 does not hold in general. The reason why the proof technique depends on strict preferences can be seen in the sketch above: we need to be able to choose the perturbations w^d to be rich enough so that orthogonality conditions (2)–(4) imply equality (5), while also ensuring that all of the functions $u_i^j \pm w^d$ still represent the ordinal type t_i.

To see that the result fails with weak preferences, revisit Example 3, but modify the preferences of agents 1 and 2 by having them be indifferent between a and a', and indifferent between b and b'. Leave the preferences of agent 3 and the SCC F as before.

Once again, F cannot be implemented by an ordinal mechanism (nothing in the original argument depended on 1's or 2's preferences between a and a' or between b and b'). However, for any interdependent cardinal extension S of T, we can implement F using the same mechanism as in the original example:

- If agent 3 has ordinal type t_3 or t_3', prescribe the $\frac{1}{2} - \frac{1}{2}$ lottery between the allowed outcomes.
- If 3's ordinal type is t_3'' and the other agents' ordinal types are not (t_1, t_2), then prescribe outcome m.
- Finally, if the agents have some interdependent cardinal types (s_1, s_2, s_3'') corresponding to ordinal types (t_1, t_2, t_3''), then choose whichever of the two lotteries $\frac{1}{2}a + \frac{1}{2}b$ or $\frac{1}{2}a' + \frac{1}{2}b'$ is preferred by the utility function $u_3(\cdot, s_1, s_2, s_3'')$.

This mechanism would not have worked with the original, strict preferences of Example 3. Under those preferences, when the ordinal type profile is (t_1, t_2, t_3''), agent 3 is supposed to get whichever of the two lotteries $\frac{1}{2}a + \frac{1}{2}b$ or $\frac{1}{2}a' + \frac{1}{2}b'$ he prefers, and this preference depends on 1's and 2's cardinal types. Hence agents 1 and 2 would have incentives to lie about their types so as to influence the choice between these two lotteries. However, in the present example, with agents 1 and 2 indifferent between a and a' and between b and b', these incentives are eliminated. ◊

The finding of this example—that the hypothesis of strict preferences in Theorem 1 cannot be abandoned—limits the theorem's applicability. As previewed in the Introduction, the strictness requirement is satisfied in voting problems such as Chatterji et al. (2014), but not in other applications such as matching. Later, in Section 5, we are able to drop the strictness requirement by relaxing the assumption of expected utility.

Example 5 (Hypothesis that F is closed-valued). The hypothesis that F should be closed-valued cannot be dispensed with. For example, let $X = \{a, b, c\}$, $n = 2$, and let there be two types of agent 1 and a single type of agent 2. Let the preferences of the agents' types and the set of lotteries allowed by F at each type profile be

	$t_2 : a \succ b \succ c$
$t_1 : a \succ b \succ c$	$\{\alpha a + (1-\alpha)c \mid 0 < \alpha < 1\}$
$t_1' : b \succ a \succ c$	$\{b\}$

This F cannot be implemented by an ordinal mechanism, since no lottery $\alpha a + (1-\alpha)c$ with $\alpha < 1$ stochastically dominates b with respect to the ordinal type t_1. But for any interdependent cardinal type space S, we can implement F over S by the mechanism that chooses $\alpha^* a + (1-\alpha^*)c$ whenever agent 1 has ordinal type t_1 and chooses b whenever 1 has ordinal type t_1', for some fixed α^*. As long as α^* is chosen sufficiently close to 1, then $\alpha^* a + (1-\alpha^*)c$ is preferred over b by the utility function $u_1(\cdot, s)$ at every cardinal type profile s with $t(s) = (t_1, t_2)$ and, hence, the incentive constraints are satisfied.

This example depends on our restriction that type spaces should be finite. Indeed, if we allow interdependent cardinal type spaces to be infinite, then we can do away with the closed-valued hypothesis; the proof is a straightforward extension of the construction used to prove Theorem 1. ◊

EXAMPLE 6 (It is not enough to consider minimal extensions). If we only assume that F is implementable over every *minimal* interdependent cardinal extension, then the conclusion of Theorem 1 does not follow. Indeed, consider once again the ordinal type space T and SCC F from Example 3. For any minimal interdependent cardinal extension S of T, we can implement F over S using the same mechanism as was given in that example: in the particular case where the type profile is (s_1, s_2, s_3''), corresponding to ordinal types (t_1, t_2, t_3''), we simply choose whichever of the two lotteries $\frac{1}{2}a + \frac{1}{2}b$ or $\frac{1}{2}a' + \frac{1}{2}b'$ is preferred by the utility function $u_3(\cdot, s_1, s_2, s_3'')$. However, as we saw in Example 3, this F cannot be implemented by an ordinal mechanism.

Thus, allowing "large" spaces of interdependent cardinal types is necessary to obtain the foundation for ordinal mechanisms. ◊

4.2 *Interpretation*

At this point, we can return to discuss modeling issues. The introduction of interdependence might seem to come out of nowhere, particularly since the original ordinal framework ((Q1) in Section 2) does not assume interdependence. We offer several arguments for why interdependence is actually natural in this setting.

- The first reason goes back to the original informal justification in Bogomolnaia and Moulin (2001) for using ordinal mechanisms: it is difficult for people to figure out their own preferences over lotteries. The natural theoretical tool to express this idea would be a model in which agents know their own ordinal preferences but face uncertainty over their cardinal preferences u_i. In such a model, each agent i actually faces *two* sources of uncertainty: his own utility function and the types of the other agents. There is no a priori reason to restrict these sources of uncertainty to be independent. Allowing for correlation leads in effect to interdependent preferences.

- A second, related argument for interdependence is practical: Virtually all real-life applications do involve common values to some extent: each agent i may have some uncertainty about fundamentals that genuinely does lead to correlation between i's preferences and other agents' information. In a voting scenario, for example, we might imagine that voter i's evaluation of each candidate x equals $\widetilde{u}_i(x) + v_i(\epsilon_x)$, where $\widetilde{u}_i(x)$ represents preferences over salient policy dimensions, ϵ_x represents less-important dimensions about which the voter is imperfectly informed, and $v_i(\epsilon_x)$ represents i's preferences over these dimensions; we assume the uncertainty in ϵ_x is small enough so as not to affect i's ordinal ranking of candidates. Voter i observes $\widetilde{u}_i(x)$ and some noisy signal $\sigma_{i,x}$ of ϵ_x. Then i's cardinal type is represented by the vector of signals σ_i, and i evaluates each candidate x by $\widetilde{u}_i(x) + \mathbb{E}[v_i(\epsilon_x) \mid \sigma_1, \ldots, \sigma_n]$; thus, we have interdependence. We could tell a similar story in school choice scenarios or other applications.

  From this point of view, the usual modeling approach, in which agents know their own preferences, simply reflects a presumption that any such interdependence is of second-order importance, so that modeling preferences as independent is a good approximation. However, we can incorporate this idea explicitly in

our framework by requiring that the amount of interdependence should be small relative to agents' certainty about their preferences, and then Theorem 1 still goes through.

Here is a formalization. Let T be an ordinal type space and let S be an interdependent cardinal extension. Fix $\epsilon > 0$. Say that S is an ϵ-*interdependent* extension if the following statements hold:

– For all $s \in S$, all agents i, and all $x, y \in X$ with $x \succ_{t_i(s_i)} y$, we have

$$u_i(x, s) > u_i(y, s) + 1.$$

– For all i and $s_i \in S_i$, and all $s_{-i}, s'_{-i} \in S_{-i}$ and all outcomes $x \in X$,

$$\left|u_i(x, s_i, s_{-i}) - u_i\left(x, s_i, s'_{-i}\right)\right| < \epsilon.$$

If we revise the hypothesis of Theorem 1 to say that F is implementable over every ϵ-interdependent cardinal extension of T, rather than every interdependent cardinal extension, the conclusion still follows. This can be shown by a straightforward, if tedious, elaboration of the original proof (we omit the details for brevity). Thus, we can interpret this version of Theorem 1 as saying that if a planner is able to implement F in a way that is robust to unknown cardinal preferences and any sufficiently small amount of cardinal interdependence, then it must be possible to achieve her goals using an ordinal mechanism.

- In addition, we point out that it is not really necessary to have the interdependence suddenly introduced at the cardinal-preference level of modeling, that is, (Q2) of Section 2. Instead, we could equally well allow for interdependence in both (Q1) and (Q2). In this case, an *ordinal type space* would consist of an abstract (finite) set $T = \times_i T_i$, together with a weak ordering $\succeq_{i,t}$ over X for each agent i at each profile $t \in T$; the incentive constraints for a ordinal mechanism $M : T \to \Delta(X)$ would then require

$$\sum_{y \succeq_{i,(t_i,t_{-i})} x} M(t_i, t_{-i})(y) \geq \sum_{y \succeq_{i,(t_i,t_{-i})} x} M\left(t'_i, t_{-i}\right)(y)$$

for all i, t_i, t'_i, and t_{-i}, and all x. The definition of an interdependent cardinal extension would then be changed simply by specifying that $u_i(\cdot, s)$ should represent the ordinal preference $\succeq_{i,t(s)}$ for each agent i and cardinal type profile s. With these definitions, Theorem 1 and its proof would go through almost unchanged.

The version where there is no interdependence at the ordinal level would then be just a special case, which happens to include the applications highlighted in the Introduction and, indeed, all of the applications studied in existing literature.

5. Non-expected utility

Ordinal mechanisms have another strength that we have so far not exploited: Not only are they robust to unknown cardinal utilities; they are robust to the possibility that

agents might not even evaluate lotteries according to expected utility theory at all. Indeed, although expected utility is orthodox in economic theory, a large body of evidence indicates that it is not always descriptive of real-world decision-makers' behavior. (See Starmer 2000 or Machina 2008 for reviews of this evidence.) To the extent that robust mechanism design aims to understand what mechanisms work well in practice, this seems an important consideration in favor of ordinal mechanisms.

To be more precise, suppose M is an ordinal mechanism implementing some SCC F. Then as long as agent i has preferences over lotteries that *respect stochastic dominance* (that is, if lottery π weakly first-order stochastically dominates π' with respect to ordering t_i, then he weakly prefers π over π'), he has no incentive to misreport his type. This is true even if these preferences over lotteries are not described by expected utility.

This suggests we should formulate (Q2) by allowing our "uncertainty about agents' preferences over lotteries" to include non-expected utility (non-EU) preferences. This is a more demanding robustness requirement than simply asking for robustness to all cardinal utility functions. We might then hope that this requirement is strong enough to imply implementation in an ordinal mechanism, thus offering an alternative foundation for ordinal mechanisms.

To investigate this then, we repeat all of our earlier analysis with a more general representation of preferences. It turns out that without interdependent preferences, the main lessons of Section 3 still go through. With interdependent preferences, however, we can give a foundation for ordinal mechanisms that is stronger than before, and we can even do so while remaining within any of several specific models of preferences that are prominent in non-EU literature.

5.1 *Non-expected utility without interdependence*

Let us first consider the case of independent preferences. A *non-EU type space*[4] is now a space $S = S_1 \times \cdots \times S_n$, where each S_i is a finite set of continuous functions $s_i : \Delta(X) \to \mathbb{R}$, used to evaluate lotteries over X. Space S is a *non-EU extension of* T if, for each i, there is a surjective function $t_i : S_i \to T_i$ such that each s_i is consistent with $t_i(s_i)$ in the sense that whenever lottery π weakly stochastically dominates π' with respect to the ordering $t_i(s_i)$, then

$$s_i(\pi) \geq s_i(\pi').$$

This is the stochastic dominance condition. It is stronger than only requiring that s_i ranks degenerate lotteries according to $t_i(s_i)$.

Let $M : S \to \Delta(X)$ be a mechanism, with $M(s) \in F(t(s))$ for all s. There are several ways to formulate the dominant-strategy incentive constraint.[5] We can say that

[4]This may be a misnomer since it includes EU preferences as a special case, but we keep this name for brevity.

[5]Existing literature on mechanism design with non-EU preferences takes varying approaches: Lopomo et al. (2014) require that there be no incentive to deviate to a mixed strategy; thus their "optimal incentive compatibility" corresponds to our "very strong implementation" below. Others (Bodoh-Creed 2012, Bose et al. 2006, Wolitzky 2016) do not consider deviations by deliberate randomization.

M weakly implements F over S if, for all i, s_i, s_i', and s_{-i},

$$s_i\big(M(s_i, s_{-i})\big) \geq s_i\big(M\big(s_i', s_{-i}\big)\big).$$

However, this only ensures that truth-telling is a dominant strategy if i is certain about the types of his opponents. If we wish to allow arbitrary uncertainty about their types, then a more appropriate constraint is that for any probability distribution σ_{-i} on S_{-i},

$$s_i\big(M(s_i, \sigma_{-i})\big) \geq s_i\big(M\big(s_i', \sigma_{-i}\big)\big),$$

where $M(s_i, \sigma_{-i})$ denotes the lottery over outcomes x that results when s_{-i} is distributed according to σ_{-i} and then x is chosen by $M(s_i, s_{-i})$ conditional on s_{-i}. If this constraint is satisfied, we say that *M strongly implements F*. With expected utility, weak implementation implies strong implementation, but with non-expected utility, this is no longer the case.

Finally, to rule out the possibility of deviations to a mixed strategy randomization by i, we can require that for all i and s_i, and all distributions σ_{-i} over S_{-i} and σ_i' over S_i,

$$s_i\big(M(s_i, \sigma_{-i})\big) \geq s_i\big(M\big(\sigma_i', \sigma_{-i}\big)\big),$$

where $M(\sigma_i', \sigma_{-i})$ is the distribution over outcomes that results from drawing type reports of i and $-i$ independently from σ_i' and σ_{-i}. If this condition is satisfied, we say that *M very strongly implements F*.

If F is implementable over T by an ordinal mechanism M, then for any non-EU extension S of T, F is very strongly implemented over S by the mechanism $s \mapsto M(t(s))$. So we are interested in the converse question: if F is implementable (in some appropriate form) over every non-EU extension S of T, must F be implementable in an ordinal mechanism? If so, we can again say that we have obtained a foundation for ordinal mechanisms.

The answer is no, for basically the same reasons as in the expected-utility case. Example 1 goes through as before as long as our concept of implementation over a non-EU extension is weak or strong implementation. With very strong implementation, the example as written does not go through: the F there is not very strongly implementable over every non-EU extension, since some type may strictly prefer the randomization $\frac{1}{4}a + \frac{1}{4}b + \frac{1}{4}c + \frac{1}{4}d$ over both $\frac{1}{2}a + \frac{1}{2}d$ and $\frac{1}{2}b + \frac{1}{2}c$. However, the example can be modified to satisfy very strong implementability as well: see Example 7 in Appendix A.

As in Section 3, we could try to restrict F to be simple. With one agent, this does give a foundation for ordinal mechanisms: Proposition 2 goes through with no trouble, since the expected-utility cardinal extensions used there are just a special case of non-EU extensions. However, with multiple agents, we again run into trouble. Example 3 suffices to show this if we use weak implementation. With strong or very strong implementation, a modification is needed. Example 8 in Appendix A shows how to do this. The SCC there can be very strongly implemented over every non-EU extension of the type space, but cannot be implemented by an ordinal mechanism. We also note that all preferences in that example are strict.

Thus the conclusions of Section 3 still hold when we allow more general preferences over lotteries.

5.2 *Non-expected utility with interdependence*

What if we allow both interdependence and general preferences over lotteries?

The first question is how to formulate preferences in this very general case. In the interdependent framework with expected utility, each type s_i of agent i had a cardinal utility defined on $X \times S_{-i}$; the outcome and the other agents' types could interact in i's preference. So the natural analogue without expected utility is to define s_i's preferences over $\Delta(X \times S_{-i})$. (We need not include S_i, since i's type is fixed from his point of view.)

Thus, we define an *interdependent non-EU type space* to be a pair (S, u), whose components are as follows:

- The set $S = S_1 \times \cdots \times S_n$ is a product of finite sets, the type spaces for each agent.
- The object u is a collection of utility functions over lotteries, one for each type s_i of each agent i, with each u_{s_i} being a continuous function $\Delta(X \times S_{-i}) \to \mathbb{R}$.

We say that (S, u) is an *interdependent non-EU extension* of T if there are surjective functions $t_i : S_i \to T_i$ such that each u_{s_i} respects stochastic dominance with respect to the weak *partial* order on $X \times S_{-i}$, given by

$$(x, s_{-i}) \succeq (y, s_{-i}) \quad \text{for all } x, y \in X \text{ such that } x \succeq_{t_i(s_i)} y \text{ and for all } s_{-i} \in S_{-i}. \tag{6}$$

When (S, u) is an interdependent non-EU extension of T and when $M : S \to \Delta(X)$ is any mechanism, we define $\widehat{M}_{-i}(s_i, s_{-i})$ for each agent i to be the distribution over $X \times S_{-i}$ that puts marginal probability 1 on type profile s_{-i} and has outcomes distributed according to $M(s_i, s_{-i})$. More generally, if σ_{-i} is a distribution over S_{-i}, we define $\widehat{M}_{-i}(s_i, \sigma_{-i})$ to be the distribution over $X \times S_{-i}$, where s_{-i} is marginally distributed according to σ_{-i} and, conditional on s_{-i}, x is distributed according to $M(s_i, s_{-i})$.

If the mechanism M satisfies $M(s) \in F(t(s))$ for all s, we can say that M *weakly implements F (in ex post equilibrium)* if, for all i, s_i, s_i', and s_{-i}, the inequality

$$u_{s_i}\big(\widehat{M}_{-i}(s_i, s_{-i})\big) \geq u_{s_i}\big(\widehat{M}_{-i}(s_i', s_{-i})\big)$$

is satisfied. Mechanism M *strongly implements F (in ex post equilibrium)* if, for all i, s_i, and s_i', and all distributions σ_{-i} over S_{-i}, we have

$$u_{s_i}\big(\widehat{M}_{-i}(s_i, \sigma_{-i})\big) \geq u_{s_i}\big(\widehat{M}_{-i}(s_i', \sigma_{-i})\big).$$

(We could further define very strong implementation, but it is not necessary to do so here.)

If we only require weak implementation over non-EU extensions, then the full preferences over lotteries of each type s_i are not needed: all that matters is the induced preference over $\Delta(X)$ for each fixed s_{-i}. In this case, Theorem 1 carries over (since EU extensions are a special case of non-EU extensions), so we have a foundation for ordinal mechanisms as long as ordinal types are strict. But we still cannot drop the assumption of strictness: Example 4 also goes through, showing that when indifferences are allowed, then weak implementation over every interdependent non-EU extension does not imply implementation in an ordinal mechanism.

However, if we strengthen our implementation concept to strong implementation, then we can give a foundation without any strictness requirement on ordinal preferences.

THEOREM 2. *Suppose $n \geq 2$, and T is any ordinal type space. Let $F : T \rightrightarrows \Delta(X)$ be an SCC with $F(t)$ closed for each $t \in T$. If F is strongly implementable over every interdependent non-EU extension of T, then it is implementable by an ordinal mechanism over T.*

(Rather than give a proof, we obtain Theorem 2 as an immediate consequence of Theorem 3 below.)

The statement of Theorem 2 is general, but somewhat unsatisfactory: it requires an extremely strong hypothesis, namely implementability for all interdependent non-EU preferences, which is a large class. We might reasonably ask what happens if the preferences are restricted to some specific class (ideally one motivated by psychological evidence), i.e., whether robustness to some particular non-EU preferences is enough for a foundation for ordinal mechanisms. There have been many such specific non-EU models studied in the literature. We content ourselves to examine a particularly prominent one here: the quadratic rank-dependent utility model.

Suppose that Y is an arbitrary finite set, that $\widetilde{u} : Y \to \mathbb{R}$ is a cardinal utility function, and that $\lambda \in [0, 1]$ is a parameter. We define a preference over $\Delta(Y)$ as follows. Order the elements of Y as $y_1, \ldots, y_r$ such that $\widetilde{u}(y_1) \leq \cdots \leq \widetilde{u}(y_r)$. Then for any $\pi \in \Delta(Y)$, define

$$u(\pi) = \sum_{j=1}^{r} \left(g_\lambda\left(\sum_{k \leq j} \pi(y_k) \right) - g_\lambda\left(\sum_{k<j} \pi(y_k) \right) \right) \widetilde{u}(y_j), \tag{7}$$

where

$$g_\lambda(p) = \lambda p^2 + (1-\lambda)\big(2p - p^2\big).$$

Thus, lotteries are evaluated with cumulative probabilities distorted according to the increasing quadratic function g_λ. (Note that if $\widetilde{u}$ assigns equal utility to several elements of Y, then there is an indeterminacy in labeling these elements, but this does not affect the value of $u(\pi)$.)

We say that $u : \Delta(Y) \to \mathbb{R}$ is a *quadratic rank-dependent utility* (QRDU) function if it has a representation of the above form for some $\widetilde{u}$ and λ. Note that any such preference over lotteries respects stochastic dominance with respect to the ordering on Y implied by $\widetilde{u}$. This form of preferences over lotteries was studied recently by Masatlioglu and Raymond (2016), who showed that it is nested in several classes of models that have figured prominently in the non-EU literature. In particular, it lies within the rank-dependent utility class of models (Quiggin 1982, Wakker 1994, Abdellaoui 2002), as well as the quadratic utility class (Chew et al. 1991). Masatlioglu and Raymond (2016) showed that with a continuous outcome space, QRDU is in fact the intersection of these two classes. Moreover, they *also* showed that it is equivalent to a case of the Kőszegi–Rabin model of reference-dependent preferences (Kőszegi and Rabin 2007), with linear gain–loss preferences and coefficient of loss aversion lying in the

interval $[0, 2]$. (The model of Kőszegi and Rabin 2007 with other values for the loss aversion coefficient would fail to respect stochastic dominance.) Finally, note that QRDU in turn nests EU, since when $\lambda = 1/2$, $g_\lambda(p) = p$ and the formula (7) collapses to expected utility.

Now returning to type spaces, say that an interdependent non-EU extension S of T is an *interdependent QRDU extension* if each function u_{s_i} is a quadratic rank-dependent utility function (on lotteries over $Y = X \times S_{-i}$). We then have the following stronger version of Theorem 2.

THEOREM 3. *Suppose $n \geq 2$ and that T is any ordinal type space. Let $F : T \rightrightarrows \Delta(X)$ be an SCC with $F(t)$ closed for each $t \in T$. If F is strongly implementable over every interdependent QRDU extension of T, then it is implementable by an ordinal mechanism over T.*

Note also that since QRDU preferences are nested within the rank-dependent utility, quadratic utility, and Kőszegi–Rabin reference-dependent preference models, Theorem 3 remains true a fortiori when QRDU is replaced by any of these three classes of preferences.

The main idea of the proof is roughly as follows. A QRDU decision-maker with parameter $\lambda > 1/2$, who compares two lotteries π and π', down-weights bad outcomes relative to an expected-utility decision-maker. Now suppose that both π and π' are diluted by a lottery π'' concentrated on medium-quality outcomes, so that the QRDU decision-maker compares $\delta\pi + (1-\delta)\pi''$ against $\delta\pi' + (1-\delta)\pi''$. As $\delta \to 0$ and $\lambda \to 1$, the bad outcomes are down-weighted so much as to become negligible relative to the good ones. Then the comparison essentially hinges on which of π or π' gives a higher probability of good outcomes. By varying π'', we can use these comparisons to replicate all of the ordinal incentive constraints.

PROOF OF THEOREM 3. Fix any small $\epsilon > 0$. We construct an interdependent QRDU extension S of T, with the property that strong implementability over S implies ϵ-implementability by an ordinal mechanism. Since this can be done for every $\epsilon > 0$, it then follows just as in the proof of Theorem 1 that F can be implemented by an ordinal mechanism.

Put $m = |X|$. Let C be a large number, specifically $C > 1/\epsilon$. Let $\gamma, \delta > 0$ be small enough so that

$$(2\epsilon - 4\delta)C^{k+1} - 4\gamma > \delta C^m + 2C^k \quad \text{for all } k = 1, \ldots, m. \tag{8}$$

This can be done, since the inequality holds when $\gamma = \delta = 0$.

These choices have the following consequence.

LEMMA 1. *If (8) holds and $q, q' \in [0, 1]$ with $q < q' - \epsilon$, then*

$$\begin{aligned}&(1-\delta q)^2 \times \left(C^{k+1} - \gamma\delta\right)\\ &\quad > \delta^2\left(1-q'\right)^2 \times C^m + \left(\left(1-\delta q'\right)^2 - \delta^2\left(1-q'\right)^2\right) \times C^{k+1} + \left(1 - \left(1-\delta q'\right)^2\right) \times C^k.\end{aligned} \tag{9}$$

The proof (a simple calculation) is provided in Appendix A.

Now we construct our extension S. Define S_i, for each agent i, to consist of one type $s_i(t_i)$ for each ordinal type $t_i \in T_i$ as well as m "dummy" types $s_i^1, \dots, s_i^m$, whose corresponding ordinal types are specified momentarily. For each t_i, rank the indifference classes of t_i from top to bottom; let $X^k[t_i]$ denote the kth indifference class from the top. Then define a cardinal preference $\widetilde{u}_{s_i(t_i)} : X \times S_{-i} \to \mathbb{R}$ as follows:

- If s_{-i} is such that, for some $k \in \{1, \dots, m\}$, we have $s_j = s_j^k$ for every agent $j \neq i$, then $\widetilde{u}_{s_i(t_i)}(x, s_{-i})$ is chosen arbitrarily in the interval $(-C^{k+1}, -C^{k+1} + \gamma\delta)$, such that $\widetilde{u}_{s_i(t_i)}(\cdot, s_{-i})$ represents the ordinal preference t_i.
- For any other s_{-i}, put $\widetilde{u}_{s_i(t_i)}(x, s_{-i}) = -C^k$ whenever $x \in X^k[t_i]$. Note that $\widetilde{u}_{s_i(t_i)}(\cdot, s_{-i})$ again represents t_i.

Now define the preference $u_{s_i(t_i)} : \Delta(X \times S_{-i}) \to \mathbb{R}$ as the QRDU preference induced by $\widetilde{u}_{s_i(t_i)}$ and parameter $\lambda = 1$. Also, for each dummy type s_i^k, arbitrarily pick an associated ordinal type $t_i \in T_i$, and let $u_{s_i^k} : \Delta(X \times S_{-i}) \to \mathbb{R}$ be any QRDU preference consistent with type t_i (for example, we could choose $u_{s_i^k}$ to be identical to $u_{s_i(t_i)}$).

This does indeed define an interdependent QRDU extension of T. In particular, the fact that preferences of $s_i(t_i)$ respect stochastic dominance with respect to ordering (6) follows from the fact that $\widetilde{u}_{s_i(t_i)}(\cdot, s_{-i})$ represents t_i for each fixed s_{-i}.

Let M be a mechanism strongly implementing F over S. We claim the ordinal mechanism M' given by $M'(t) = M(s(t))$ must ϵ-implement F over T. Certainly $M'(t) \in F(t)$, so we just need to check the ordinal incentive constraints for ϵ-implementation.

Consider any t_i, t_i', and t_{-i}, and any $k = 1, \dots, m$. Consider the probability distribution σ_{-i} over S_{-i} that puts weight δ on $s_{-i}(t_{-i})$ and puts remaining weight $1 - \delta$ on s_{-i}^k (defined by taking dummy type s_j^k for each agent $j \neq i$). The strong incentive constraint for M, with this probability distribution over s_{-i}, tells us that

$$\begin{aligned} &u_{s_i(t_i)}\big(\delta \widehat{M}_{-i}\big(s_i(t_i), s_{-i}(t_{-i})\big) + (1-\delta)\widehat{M}_{-i}\big(s_i(t_i), s_{-i}^k\big)\big) \\ &\quad \geq u_{s_i(t_i)}\big(\delta \widehat{M}_{-i}\big(s_i\big(t_i'\big), s_{-i}(t_{-i})\big) + (1-\delta)\widehat{M}_{-i}\big(s_i\big(t_i'\big), s_{-i}^k\big)\big). \end{aligned} \tag{10}$$

Let q be the total probability placed by lottery $M(s_i(t_i), s_{-i}(t_{-i}))$ on the top k indifference classes of t_i; therefore, $1 - q$ is the total probability placed on lower indifference classes. (If k exceeds the number of indifference classes, then $q = 1$.) The left-hand side of (10) evaluates a lottery on $X \times S_{-i}$ that, by construction, has the following properties:

(i) It places probability $\delta(1 - q)$ on outcomes with cardinal utility in the interval $[-C^m, -C^{k+1}]$.

(ii) It places probability $1 - \delta$ on outcomes with cardinal utility in $(-C^{k+1}, -C^{k+1} + \gamma\delta)$.

(iii) It places δq on outcomes with cardinal utility in $[-C^k, 0]$.

To evaluate this lottery according to (7), outcomes are weighted so that the cumulative probabilities are distorted by $g_1(p) = p^2$. Then the outcomes in (i) and (ii) receive total

weight $(1-\delta q)^2$, and those in (iii) receive the remaining weight $1-(1-\delta q)^2$. Therefore, the value of this lottery according to $u_{s_i(t_i)}$ is at most $(1-\delta q)^2 \times (-C^{k+1}+\gamma\delta)$.

Similarly, let q' be the total probability placed by lottery $M(s_i(t_i'), s_{-i}(t_{-i}))$ on the top k indifference classes of t_i. The right-hand side of (10) evaluates a lottery that has the following properties:

(i′) It places probability $\delta(1-q')$ on outcomes with cardinal utility in $[-C^m, -C^{k+1}]$.

(ii′) It places probability $1-\delta$ on outcomes with cardinal utility in $(-C^{k+1}, -C^{k+1}+\gamma\delta)$.

(iii′) It places $\delta q'$ on outcomes with cardinal utility in $[-C^k, 0]$.

Weighting according to g_1, the value of this lottery according to $u_{s_i(t_i)}$ is at least

$$\delta^2(1-q')^2 \times (-C^m) + \left((1-\delta q')^2 - \delta^2(1-q')^2\right) \times (-C^{k+1}) + \left(1-(1-\delta q')^2\right) \times (-C^k).$$

Putting these bounds together with (10), we get

$$\begin{aligned}
&(1-\delta q)^2 \times (-C^{k+1}+\gamma\delta) \\
&\quad \geq \delta^2(1-q')^2 \times (-C^m) + \left((1-\delta q')^2 - \delta^2(1-q')^2\right) \times (-C^{k+1}) \\
&\qquad + \left(1-(1-\delta q')^2\right) \times (-C^k).
\end{aligned} \tag{11}$$

Now if $q < q' - \epsilon$, then Lemma 1 tells us that (9) holds. But (9) is exactly the negation of (11) (after multiplying through by -1); thus we get a contradiction. We conclude that $q \geq q' - \epsilon$.

At this point, we have shown that the total probability placed on the top k indifference classes of t_i in lottery $M(s_i(t_i'), s_{-i}(t_{-i}))$ cannot exceed the corresponding probability in $M(s_i(t_i), s_{-i}(t_{-i}))$ by more than ϵ. Since this holds for each k, this exactly gives the ordinal incentive constraints for M' to ϵ-implement F.

Now we take limits as $\epsilon \to 0$ to infer that F can be implemented by an ordinal mechanism, just as in the proof of Theorem 1. □

A couple of remarks are in order. First, we have used only QRDU with parameter $\lambda = 1$ here. We could also have instead used $\lambda = 0$, so that agents underweight good outcomes rather than bad ones, and a parallel argument would apply.

However, we could not restrict to λ bounded strictly away from 0 and 1, and expect the same proof to work. That is, the proof technique would not succeed if we only allowed QRDU preferences that are "close" to EU. We also could not restrict the interdependence in preferences to be "small" (as we could in Theorem 1; see Section 4.2). While the proof above would not work under either of these restrictions, whether Theorem 3 would remain true is an open question.

6. Other extensions

There are various ways one can elaborate on the preceding ideas. We indicate a couple of such extensions.

6.1 *Other sets of utility functions*

We have taken as the primitives a space of ordinal types T_i for each agent and an SCC $F : T \to \Delta(X)$, and looked for a theoretical foundation for restricting attention to mechanisms in which agents directly report their ordinal type t_i. Given that T is the domain of F, we can think of it as encapsulating the information about agents' preferences that is relevant to the planner's goals. Under interdependence, T also simultaneously represents the information that each agent is presupposed to know for sure about his own preferences.

We have used ordinal preferences as a primitive because of the existing mechanism design literature in this framework. However, we could also imagine allowing the planner's goals to depend on preference information at some other level of granularity, e.g., only on each agent's k most preferred outcomes, or on preferences between some small set of lotteries. We could again ask if such a planner is justified in restricting attention to mechanisms that only elicit this information, based on uncertainty about finer preferences.

For this more general model, following the ideas of Carroll (2010, 2012), Dubra et al. (2004), we could define a *type* t_i of agent i to be a nonempty set of utility functions from X to $\mathbb{R}$. A *type space* would be a finite set of types. (Ordinal types then consist of sets of the form $\{u_i \mid u_i \text{ represents } \succeq_i\}$ for some weak order $\succeq_i$ on X.) Cardinal extensions would be defined as before, with the requirement "s_i represents t_i" changed to "$s_i \in t_i$," and analogously for interdependent cardinal extensions. A *mechanism* $M : T \to \Delta(X)$ would *implement* F if $M(t) \in F(t)$ for all t, and $u_i(M(t_i, t_{-i})) \geq u_i(M(t_i', t_{-i}))$ for all t_i, t_i', t_{-i} and all $u_i \in t_i$.

Theorem 1 can be generalized to this setting. Instead of imposing that preferences are strict, the requirement would be that each type t_i of each agent should be an open set. The proof requires only minor adaptations.

6.2 *Bayesian implementation*

All of our analysis has been based on ex post implementation. This is in accordance with the existing literature, as discussed in the Introduction. However, it is natural to try to ask the same questions in the paradigm of Bayesian implementation instead. We briefly present findings here; details of the definitions and examples can be found in Appendix B. We assume $n \geq 2$, since Bayesian and dominant-strategy implementation coincide for $n = 1$.

To study Bayesian implementation, type spaces (in either their ordinal or cardinal versions) need to be supplemented with priors: each type of each agent should be endowed with a distribution over the types of other agents. Then, in an ordinal environment, the natural incentive-compatibility condition for a direct mechanism is *ordinal Bayesian incentive compatibility* (OBIC) (d'Aspremont and Peleg 1988, Majumdar and Sen 2004, Bhargava et al. 2015, Mishra 2016). This criterion says that for each agent i, the lottery he gets by reporting his true type t_i always first-order stochastically dominates any lottery he could get by reporting another type t_i', where the lotteries

result from i's subjective uncertainty about others' types as well as any possible randomization in the mechanism conditional on the type profile.

Now consider first cardinal extensions without interdependence. If an SCC F can be implemented by an ordinal mechanism (using ordinal Bayesian incentive compatibility as the criterion), then it is also implementable over any cardinal extension of the type space, with compatible priors. If the converse were true, we would have a foundation for ordinal mechanisms in the Bayesian framework. Unfortunately, this is not the case. The Appendix shows a simple counterexample, extending the ideas from Examples 1 and 3.

What happens when we allow interdependence? In this case, even the "forward" direction, from (Q1) to (Q2), fails: an SCC may be implemented (in the OBIC sense) by an ordinal mechanism, yet not implementable over an interdependent cardinal extension of the type space. The Appendix illustrates this with a simple example. So asking the converse question seems unmotivated. That is, in the Bayesian setting with interdependence, it is not even clear what statement one would try to prove to give a foundation for ordinal mechanisms.

7. Summary

We close with a review of our main findings and a brief discussion.

This paper was motivated by recent literature on randomized mechanisms for agents with ordinal preferences over outcomes. This literature generally looks for mechanisms in which agents report their ordinal preferences, and truthfulness is a dominant strategy. We undertook a quest for theoretical justifications for restricting attention to ordinal mechanisms. Specifically, following the approach of Bergemann and Morris (2005), we asked the following question. Suppose the agents have preferences over lotteries, but these preferences are not known to the planner, and the planner's goals do not depend on more than the ordinal preferences. Suppose that these goals can always be implemented by *some* mechanism no matter what the agents' true preferences are (with ex post implementation as the solution concept). Does it follow that the goals can be implemented by an ordinal mechanism? If so, we say that we have a *foundation* for ordinal mechanisms. In this case, the desire for robustness to uncertainty about the agents' exact preferences allows the planner to restrict her attention to ordinal mechanisms.

Whether this foundation exists depends on just how much robustness is desired. There are several robustness criteria that do give such a foundation.

- If there is just one agent, then there is a foundation for ordinal mechanisms if we assume the SCC expressing the planner's goals is simple (Proposition 2).
- If there are multiple agents whose ordinal preferences over all outcomes are strict, and the planner desires robustness to interdependence in their cardinal preferences, then we again obtain a foundation for ordinal mechanisms (Theorem 1). Interdependence can be justified by supposing that agents have some uncertainty about their own cardinal preferences, which might depend on unobserved fundamentals and thereby be correlated with other agents' types. The result holds even if we require robustness only to a small amount of interdependence.

- If there are multiple agents and weak preferences are possible, but the planner desires robustness to interdependence and to non-expected utility, then we again have a foundation for ordinal mechanisms (Theorem 2). This foundation remains valid if preferences are restricted to the QRDU class (Theorem 3), which lies within several commonly studied models: rank-dependent utility, quadratic utility, and Kőszegi–Rabin reference-dependent preferences.

However, in each of these cases, if we remove any one of the conditions, then we lose the foundation for ordinal mechanisms. Specifically, in each of the following situations, we can give an SCC that is robustly implementable but not using an ordinal mechanism:

- One agent and a non-simple SCC (Example 1).
- Multiple agents who know their own preferences over lotteries, even if it is required that ordinal preferences are strict, the SCC is simple and arbitrary non-expected utility is possible (Examples 3 and 8).
- Multiple agents with weak ordinal preferences and interdependent cardinal preferences, but who adhere to expected utility (Example 4).

We have adopted a modeling framework that focuses on possibility or impossibility of implementing a given SCC. Not all questions that have been studied in the ordinal mechanism literature fall within this framework. In particular, results that characterize all mechanisms that implement an SCC, such as the random dictatorship result of Chatterji et al. (2014), do not map exactly into our framework. However, the approach here seems to be natural if we want a framework that transcends any particular application.

Our results can be interpreted positively or negatively. The positive interpretation is that they provide a justification for looking at ordinal mechanisms in situations where a sufficient amount of robustness is desired. For example, return to the object allocation problem of Bogomolnaia and Moulin (2001) from the Introduction. They showed that no strategy-proof ordinal mechanism guarantees an ordinally efficient random allocation that satisfies equal treatment of equals. We can apply Theorem 2 to conclude that *no* ex-post mechanism, ordinal or not, guarantees these properties while being robust to interdependent, non-expected-utility preferences.

The negative interpretation of our findings is that if the planner does not desire too much robustness, then focusing attention on ordinal mechanisms may entail a loss of generality. From this point of view, a fully satisfactory analysis of such a mechanism design problem should allow agents to express non-ordinal preferences.

Unfortunately, the study of strategy-proof randomized mechanisms that elicit cardinal preferences has proven to be analytically quite difficult. (The literature on such problems consists of a small handful of papers; see Zhou 1990, Barberà et al. 1998, Freixas 1984, Hylland 1980, Schummer 1999, Filos-Ratsikas et al. 2014.) For this reason, realistically, future theoretical work in these domains is likely to continue focusing on ordinal mechanisms.

One potential direction for progress, rather than focusing on impossibility and possibility results as in this paper, would be to identify situations where a small departure

from ordinal mechanisms provides clear benefits. Here is a simple example, based on the idea of Example 1.

Return to the object allocation setting as in Bogomolnaia and Moulin (2001), with four agents, but now suppose that all agents are known in advance to have the same ranking over objects, $a \succ b \succ c \succ d$. Since we are now interested in making low robustness demands, assume agents know their own cardinal preferences.

If a mechanism can only use ordinal preferences (which are already known), then essentially the only fair allocation would be assigning the objects uniformly at random; let π denote this lottery. But as an alternative, consider the following non-ordinal mechanism. Take the lottery π' that allocates a and d randomly (uniformly) between the first two agents, and allocates b and c randomly between the last two. Ask each agent to report a preference between π and π', and choose π' only if all agents prefer it. This mechanism retains the advantage of strategy-proofness and gives a (cardinal) Pareto improvement over the ordinal mechanism that simply assigns π.

Looking for easy improvements of this sort could be a natural direction to advance beyond purely ordinal mechanisms in specific applications. (Kesten 2010 makes a somewhat related proposal in a school choice setting.)

Appendix A: Omitted proofs and examples

Proof of Proposition 1. Part (a). Suppose F is an SCF that cannot be implemented by an ordinal mechanism. So for some i, there exist t_i, t_i', and t_{-i} such that $F(t_i, t_{-i})$ does not first-order stochastically dominate $F(t_i', t_{-i})$ with respect to preferences t_i. Then we can find a cardinal utility function s_i that represents t_i such that $s_i(F(t_i, t_{-i})) < s_i(F(t_i', t_{-i}))$. Take S to be any cardinal extension of T in which s_i is a possible type of agent i; then F cannot be implemented over S.

Part (b). Suppose F is deterministic. Let S be any cardinal extension of T, and suppose F is implemented over S by mechanism M. For each $t_i \in T$, choose some $s_i(t_i) \in S_i$ that represents t_i. Write $s(t) = (s_1(t_1), \dots, s_n(t_n))$. Then F is implemented over T by the ordinal mechanism $t \mapsto M(s(t))$. (In effect, comparison by expected utility is the same as comparison by stochastic dominance when the lotteries being compared are degenerate.) □

Proof of Proposition 2. For every ordinal type t_1 and every nonempty set $Z \subseteq X$, write $\mathrm{Top}(Z|t_1)$ for the set of elements of Z that are most preferred by type t_1. (This set may contain more than one element if there are indifferences.) Also, for each ordinal type $t_1 \in T_1$, define $G(t_1)$ to be the nonempty subset of X such that $F(t_1) = \Delta(G(t_1))$.

We now construct a sequence of subsets $Z^0, \dots, Z^r \subseteq X$ and a sequence of types $t_1^1, \dots, t_1^r \in T_1$ as follows.

- Step 0. Set $Z^0 = X$.
- Step $k > 0$. Suppose $Z^0, \dots, Z^{k-1}$ have been defined so far.
  - If $Z^{k-1} = \varnothing$, then stop, setting $r = k - 1$.

– If, for every t_1, we have $G(t_1) \cap \text{Top}(Z^{k-1}|t_1) \neq \varnothing$, then stop, setting $r = k-1$.

– Otherwise, choose some t_1^k such that $G(t_1^k) \cap \text{Top}(Z^{k-1}|t_1^k) = \varnothing$. Let $Z^k = Z^{k-1} \setminus \text{Top}(Z^{k-1}|t_1^k)$.

Since $Z^0, Z^1, \ldots$ is a strictly decreasing sequence of subsets of the finite set X, the algorithm must eventually terminate.

The interpretation of the algorithm is that we successively eliminate elements of X that cannot be chosen with positive probability by $M(t_1)$ for any t_1, if M is to be an ordinal mechanism that implements F. To understand this, consider Step 1 and suppose some type t_1^1 is chosen in that step. If the desired mechanism M exists, it cannot give type t_1^1 any of its top-ranked outcomes (in X) with positive probability, since the SCC prohibits them. Then ordinal incentive compatibility implies that none of t_1^1's top-ranked outcomes can be assigned positive probability under *any* type t_1; so these outcomes are eliminated entirely from the range of M. Thus, for all t_1, we have $M(t_1) \in \Delta(Z^1)$. This argument can be iterated to show that $M(t_1) \in \Delta(Z^k)$ for each k.

Now, if the final set Z^r is nonempty, then we can find a deterministic, ordinal mechanism M that implements F. Namely, for each t_1, let $M(t_1)$ put probability 1 on some outcome in $G(t_1) \cap \text{Top}(Z^r|t_1)$. (Such an outcome exists, by the termination condition for the algorithm.) Then each type t_1 is assigned one of its favorite elements of the set Z^r, and since agent 1 could only get a different element of Z^r by misreporting, each type's incentive constraints are satisfied. Moreover, M assigns to each type t_1 one of the acceptable outcomes in $G(t_1)$. So M implements F.

Therefore, if F cannot be implemented by such a mechanism, then $Z^r = \varnothing$. In this case, we proceed to construct a minimal cardinal extension of T_1 over which F cannot be implemented. Let C be a large positive constant.

We construct, for each type $t_1 \in T_1$, a utility function $s_1(t_1)$ as follows.

- If t_1 is not equal to t_1^k for any k, then s_1 may be any utility function that represents t_1.

- Otherwise, for each $x \in X$, define its *rank* with respect to t_1, notated $k(x)$ (with the dependence on t_1 implicit), to be the earliest step k such that $t_1 = t_1^k$ and $x \succeq_{t_1} \text{Top}(Z^{k-1}|t_1)$ (this abuse of notation is unambiguous since t_1 is indifferent among all elements of $\text{Top}(Z^{k-1}|t_1)$). If no such step exists, then take $k(x) = r+1$.

  Notice that if $x \succeq_{t_1} y$, then $k(x) \leq k(y)$; hence, $C^{3^{r+1-k(x)}} \geq C^{3^{r+1-k(y)}}$. Consequently, we can find a utility function $s_1(t_1)$ that represents t_1, such that

$$s_1(t_1)(x) \in \left[C^{3^{r+1-k(x)}}, C^{3^{r+1-k(x)}} + 1\right]$$

  for all outcomes x.

Thus each $s_1(t_1)$ represents t_1.

Let S_1 be the space consisting of the cardinal types thus constructed. Suppose, seeking a contradiction, that some mechanism M represents F over S_1.

We claim that for each k and each type s_1, $M(s_1)$ puts total probability at most $1/C^{3^{r-k}}$ on outcomes not in Z^k.

The proof of the claim is by induction. The base case $k=0$ is trivial since $Z^0 = X$. If the claim holds for $k-1$, then consider the ordinal type $t_1^* = t_1^k$ chosen at step k, with its corresponding cardinal type s_1^*. This ordinal type may have been chosen at some previous steps as well. Thus, let $k_1 < k_2 < \cdots < k_q = k$ be all the steps $k' \le k$ such that $t_1^* = t_1^{k'}$ (we may have $q=1$ if there were no previous such steps). Also put $k_0 = 0$ for convenience.

We now bound from above the utility attained by type s_1^* in the mechanism M. Consider any outcome x of rank $\le k$ with respect to t_1^*. This rank must be k_j for some j. By definition of rank, $x \succeq_{t_1^*} \mathrm{Top}(Z^{k_j-1}|t_1^*)$. Now if x is assigned positive probability by $M(s_1^*)$, then $x \notin \mathrm{Top}(Z^{k_j-1}|t_1^*)$, since by construction the latter set consists only of outcomes that are not in $G(t_1^{k_j}) = G(t_1^*)$. Therefore, $x \notin Z^{k_j-1}$. Since $k_j - 1 < k$, we can apply the induction hypothesis to see that the set of all x of rank k_j then has total probability at most $1/C^{3^{r-(k_j-1)}}$.

However, any x of rank k_j gives cardinal utility at most $C^{3^{r+1-k_j}} + 1$ to type s_1^*, by virtue of the construction of its utility function. So the outcomes of rank k_j contribute a total of at most

$$\frac{C^{3^{r+1-k_j}} + 1}{C^{3^{r-(k_j-1)}}} = 1 + C^{-3^{r+1-k_j}}$$

to the utility of type s_1^*. The right-hand side is certainly smaller than, say, 2. Finally, summing over all of the k_j, we see that the outcomes of rank at most k contribute at most $2r$ to the utility of type s_1^*.

Meanwhile, the outcomes of any rank at least $k+1$ each give type s_1^* a utility of at most $C^{3^{r+1-(k+1)}} + 1 = C^{3^{r-k}} + 1$. So altogether, the utility achieved by type s_1^* in mechanism M satisfies

$$s_1^*\big(M(s_1^*)\big) \le C^{3^{r-k}} + 2r + 1.$$

However, every outcome in $Z^{k-1} \setminus Z^k = \mathrm{Top}(Z^{k-1}|t_1^*)$ has rank k and so gives s_1^* a cardinal utility of at least $C^{3^{r+1-k}}$. Hence, for any cardinal type s_1, the lottery $M(s_1)$ can assign total probability at most

$$\frac{C^{3^{r-k}} + 2r + 1}{C^{3^{r+1-k}}} = C^{(-2)\cdot 3^{r-k}} + (2r+1)C^{-3^{r+1-k}}$$

to these outcomes, otherwise s_1^* would benefit from imitating s_1 in mechanism M, violating the incentive constraint.

Combining this with the induction hypothesis—that for all cardinal types s_1, the lottery $M(s_1)$ assigns total probability at most $C^{-3^{r+1-k}}$ to outcomes not in Z^{k-1}—we

conclude that $M(s_1)$ assigns total probability at most

$$C^{-3^{r+1-k}} + C^{(-2)\cdot 3^{r-k}} + (2r+1)C^{-3^{r+1-k}}$$

to outcomes not in Z^k. As long as C is large enough, this expression is less than $C^{-3^{r-k}}$. Thus, we can see that for all s_1, the total probability assigned by lottery $M(s_1)$ to outcomes not in Z^k is at most $C^{-3^{r-k}}$. This completes the induction step and proves the claim.

Finally, taking $r = k$ in the claim, we see that for each s_1, the total probability assigned to outcomes in $X \setminus Z^r$ is at most $C^{-3^{r-r}} = C^{-1}$. But $Z^r = \varnothing$, so this total probability should be 1, a contradiction.

So either our elimination algorithm leads to $Z^r \neq \varnothing$, in which case F can be implemented by a deterministic ordinal mechanism, or leads to $Z^r = \varnothing$, in which case we have a minimal cardinal extension of T_1 over which F is not implementable. The proposition follows. □

Proof of Theorem 1. Fix any $\epsilon > 0$. Define ϵ-implementation by an ordinal mechanism as in the text. We construct an interdependent cardinal extension S of T, such that if F is implementable over S, then F is ϵ-implementable by an ordinal mechanism.

Put $m = |X|$ and label the outcomes as $X = \{x_1, \dots, x_m\}$. As in the sketch in the text, we identify utility functions with vectors in $\mathbb{R}^m$, so that expected utility for a lottery is given by the inner product. Also, let K be a positive integer divisible by all of the numbers $1, 2, \dots, 2m$.

Now, for each agent i, and each ordinal type t_i, we define $K(m-1)+1$ interdependent cardinal types

$$\begin{aligned}
&s_{t_i}^{1,0}, s_{t_i}^{1,1}, \dots, s_{t_i}^{1,K-1},\\
&s_{t_i}^{2,0}, s_{t_i}^{2,1}, \dots, s_{t_i}^{2,K-1},\\
&\vdots\\
&s_{t_i}^{m-1,0}, s_{t_i}^{m-1,1}, \dots, s_{t_i}^{m-1,K-1},\\
&s_{t_i}^{m,0}.
\end{aligned}$$

Let S_i consist of the $|T_i| \cdot (K(m-1)+1)$ symbols thus defined and let $S = S_1 \times \cdots \times S_n$. Of course, when we define S as an extension of T below, each of the types in S_i created above is associated with ordinal type t_i. The types $s_{t_i}^{j,0}$ serve as the endpoints of the staircases, analogous to s_i^1 and s_i^7 in Figure 2, and $s_{t_i}^{j,k}$ for $k > 0$ denote the intermediate types along the staircase from $s_{t_i}^{j,0}$ to $s_{t_i}^{j+1,0}$.

It is helpful to give short names to the various type profiles that form the staircases in the construction below. First, for each $j = 1, \dots, m-1$, define $s_{t_i}^{j,K} = s_{t_i}^{j+1,0}$ for convenience. For any $t \in T$ and any $(j,k) \in \{1, \dots, m-1\} \times \{0, \dots, K\}$ or $(j,k) = (m,0)$, write

$$s_t^{j,k} = \left(s_{t_1}^{j,k}, s_{t_2}^{j,k}, \dots, s_{t_n}^{j,k}\right).$$

For each value $d \in \{1, \dots, 2m\}$, and each $j \in \{1, \dots, m-1\}$ and $k \in \{0, 1, \dots, K-d\}$, define the profiles

$$s_t^{j,k,d,i\to} = \left(s_{t_1}^{j,k}, s_{t_2}^{j,k}, \dots, s_{t_{i-1}}^{j,k}, s_{t_i}^{j,k+d}, s_{t_{i+1}}^{j,k+d}, \dots, s_{t_n}^{j,k+d}\right)$$

for all agents $i = 2, \dots, n$.

Thus, the profiles $s_t^{j,k}$ are those in which all agents' cardinal types have the same superscript (j,k), and the profiles $s_t^{j,k,d,i\to}$ are those for which agents from i onward have superscript $(j, k+d)$ and earlier agents have superscript (j,k). Notice that all the profiles of these forms are distinct, aside from the identities $s_t^{j,K} = s_t^{j+1,0}$.

We next specify agents' utility functions along the staircases. For each agent i, each ordinal type t_i, and each outcome x_j, define a function $u_{t_i}^{j,0} : X \to \mathbb{R}$ that represents t_i and such that

$$u_{t_i}^{j,0}(x) \in \left[1 - \frac{\epsilon}{2}, 1\right] \quad \text{if } x \succeq_{t_i} x_j,$$

$$u_{t_i}^{j,0}(x) \in \left[0, \frac{\epsilon}{2}\right] \quad \text{otherwise.}$$

Clearly this can be done. For each $j = 1, \dots, m-1$ and each $k = 1, \dots, K-1$, let $u_{t_i}^{j,k} : X \to \mathbb{R}$ be any arbitrary utility function that represents t_i.

Also, let $w^1, w^2, \dots, w^m$ be any basis for the linear space of utility functions, $\mathbb{R}^m$. By scaling, assume that all the w^d are chosen close enough to 0 so that all functions of the form $u_{t_i}^{j,k} \pm w^d$ still represent t_i for each agent i and ordinal type t_i. Additionally, define $w^{m+1}, w^{m+2}, \dots, w^{2m}$ by $w^{m+d} = -w^d$.

Now we specify the utility functions $u_i(x,s)$. For all $t \in T$ and all $x \in X$, let

$$u_i\left(x, s_t^{j,k}\right) = u_{t_i}^{j,k}(x)$$

for all agents i and all (j,k), and let

$$u_i\left(x, s_t^{j,k,d,i+1\to}\right) = u_{t_i}^{j,k+d}(x) - w^d(x),$$

$$u_i\left(x, s_t^{j,k,d,i\to}\right) = u_{t_i}^{j,k+d}(x), \qquad \text{if } i < n$$

$$u_n\left(x, s_t^{j,k,d,n\to}\right) = u_{t_n}^{j,k}(x) + w^d(x),$$

each for all (j,k,d,i) for which the relevant cardinal type profiles are defined. (These definitions are to be made for all $t \in T$ and all $x \in X$. Here we take $u_{t_i}^{j,K} = u_{t_i}^{j+1,0}$.)

There are no inconsistencies in the specifications we have made so far; that is, we have never defined the utility function for the same agent at the same type profile twice. Moreover, for each agent i, all the utility functions we have assigned at any cardinal type profile in which i has a type $s_{t_i}^{j,k}$ do indeed represent the ordinal type t_i.

Finally, for each agent i, for all of the cardinal type profiles s for which we have not yet defined a utility function, we can simply let $u_i(\cdot, s)$ be any function at all that represents the ordinal type associated with s_i. This completes the construction of (S,u) and ensures that it is indeed an interdependent cardinal extension of T.

Now let M be any mechanism that implements F in ex post equilibrium over (S, u). Consider any fixed $t \in T$. We show the "staircase equalities"

$$M\big(s_t^{1,0}\big) = M\big(s_t^{2,0}\big) = \cdots = M\big(s_t^{m,0}\big). \tag{12}$$

To this end, fix $j \in \{1, \ldots, m-1\}$. Consider any $d \in \{1, \ldots, 2m\}$ and $k \in \{0, \ldots, K-d\}$.

The two type profiles $s_t^{j,k}$ and $s_t^{j,k,d,n\rightarrow}$ differ only in the types of agent n: $s_{t_n}^{j,k}$ in the former and $s_{t_n}^{j,k+d}$ in the latter. Thus, the incentive constraints for agent n at the two profiles give us

$$u_{t_n}^{j,k} \cdot M\big(s_t^{j,k}\big) \geq u_{t_n}^{j,k} \cdot M\big(s_t^{j,k,d,n\rightarrow}\big)$$

and

$$\big(u_{t_n}^{j,k} + w^d\big) \cdot M\big(s_t^{j,k,d,n\rightarrow}\big) \geq \big(u_{t_n}^{j,k} + w^d\big) \cdot M\big(s_t^{j,k}\big).$$

Subtracting these two inequalities gives

$$w^d \cdot \big(M\big(s_t^{j,k,d,n\rightarrow}\big) - M\big(s_t^{j,k}\big)\big) \geq 0. \tag{13}$$

For each $i = 2, \ldots, n-1$, the two type profiles $s_t^{j,k,d,i+1\rightarrow}$ and $s_t^{j,k,d,i\rightarrow}$ differ only in the types of agent i. The incentive constraints for agent i give

$$\big(u_{t_i}^{j,k+d} - w^d\big) \cdot M\big(s_t^{j,k,d,i+1\rightarrow}\big) \geq \big(u_{t_i}^{j,k+d} - w^d\big) \cdot M\big(s_t^{j,k,d,i\rightarrow}\big)$$

and

$$u_{t_i}^{j,k+d} \cdot M\big(s_t^{j,k,d,i\rightarrow}\big) \geq u_{t_i}^{j,k+d} \cdot M\big(s_t^{j,k,d,i+1\rightarrow}\big).$$

Subtracting gives

$$w^d \cdot \big(M\big(s_t^{j,k,d,i\rightarrow}\big) - M\big(s_t^{j,k,d,i+1\rightarrow}\big)\big) \geq 0. \tag{14}$$

Finally, when $i = 1$, the two type profiles $s_t^{j,k,d,2\rightarrow}$ and $s_t^{j,k+d}$ again differ only in the types of agent 1, and the incentive constraints give

$$\big(u_{t_1}^{j,k+d} - w^d\big) \cdot M\big(s_t^{j,k,d,2\rightarrow}\big) \geq \big(u_{t_1}^{j,k+d} - w^d\big) \cdot M\big(s_t^{j,k+d}\big)$$

and

$$u_{t_1}^{j,k+d} \cdot M\big(s_t^{j,k+d}\big) \geq u_{t_1}^{j,k+d} \cdot M\big(s_t^{j,k,d,2\rightarrow}\big),$$

from which we subtract to obtain

$$w^d \cdot \big(M\big(s_t^{j,k+d}\big) - M\big(s_t^{j,k,d,2\rightarrow}\big)\big) \geq 0. \tag{15}$$

Combining (13), (14), and (15) now gives

$$w^d \cdot \big(M\big(s_t^{j,k+d}\big) - M\big(s_t^{j,k}\big)\big) \geq 0. \tag{16}$$

Now, whenever $k' \in \{1, \dots, K\}$ such that $k' - k$ is a positive multiple of d, we can apply (16) to the numbers $k, k+d, k+2d, \dots, k'-d$, and combine to obtain

$$w^d \cdot \big(M\big(s_t^{j,k'}\big) - M\big(s_t^{j,k}\big)\big) \geq 0.$$

In particular, for any $d = 1, 2, \dots, 2m$, we can choose $k' = K$ and $k = 0$, and we get

$$w^d \cdot \big(M\big(s_t^{j,K}\big) - M\big(s_t^{j,0}\big)\big) \geq 0. \tag{17}$$

Now, for each $d = 1, \dots, m$, (17) holds with w^d and also holds with $w^{d+m} = -w^d$; hence, we get

$$w^d \cdot \big(M\big(s_t^{j,K}\big) - M\big(s_t^{j,0}\big)\big) = 0.$$

Since the vector $M(s_t^{j,K}) - M(s_t^{j,0})$ is orthogonal to all of the w^d, which span the space $\mathbb{R}^m$, it must be zero. Thus $M(s_t^{j,K}) = M(s_t^{j,0})$. In view of $s_t^{j,K} = s_t^{j+1,0}$, we conclude that (12) holds.

Now define an ordinal mechanism $M' : T \to \Delta(X)$ by

$$M'(t) = M\big(s_t^{j,0}\big).$$

In view of (12), the definition is the same regardless of which j we choose. We claim that the resulting mechanism M' ϵ-implements F over T. We have $M'(t) = M(s_t^{j,0}) \in F(t(s_t^{j,0})) = F(t)$ (for any j), so we need only to check the ϵ-incentive constraint. Consider any i, t_i, t_i', and t_{-i}. Choose any outcome in X, say x_j. For any lottery π, we have

$$\sum_{y \succeq_{t_i} x_j} \pi(y) - \frac{\epsilon}{2} \leq u_{t_i}^{j,0}(\pi) \leq \frac{\epsilon}{2} + \sum_{y \succeq_{t_i} x_j} \pi(y)$$

by construction of $u_{t_i}^{j,0}$. Therefore,

$$\begin{aligned}
\sum_{y \succeq_{t_i} x_j} M'(t)(y) &\geq u_{t_i}^{j,0}\big(M'(t)\big) - \frac{\epsilon}{2} \\
&= u_i\big(M'(t), s_t^{j,0}\big) - \frac{\epsilon}{2} \\
&\geq u_i\big(M'\big(t_i', t_{-i}\big), s_t^{j,0}\big) - \frac{\epsilon}{2} \\
&= u_{t_i}^{j,0}\big(M'\big(t_i', t_{-i}\big)\big) - \frac{\epsilon}{2} \\
&\geq \sum_{y \succeq_{t_i} x_j} M'\big(t_i', t_{-i}\big)(y) - \frac{\epsilon}{2} - \frac{\epsilon}{2}.
\end{aligned}$$

(Here the middle inequality is exactly the incentive constraint of type $s_{t_i}^{j,0}$ in mechanism M at profile $s_t^{j,0}$, stating that agent i does not wish to misreport as type $s_{t_i'}^{j,0}$.) Thus the ϵ-incentive constraint (1) is verified.

This shows that F can be ϵ-implemented by an ordinal mechanism over T.

Let us complete the proof. For each $\epsilon > 0$, we can find an $M^\epsilon : T \to \Delta(X)$ that ϵ-implements F. By using compactness and passing to a subsequence if necessary, we have a sequence of values $\epsilon \to 0$ along which the $M^\epsilon(t)$ converge for every t. Let $M^0(t)$ be the corresponding limit. This defines an ordinal mechanism M^0. We have $M^0(t) \in F(t)$ for each t because $M^\epsilon(t) \in F(t)$ and $F(t)$ is closed. Moreover, for each ϵ, the mechanism M^ϵ satisfies every incentive constraint (1); taking limits as $\epsilon \to 0$, we see that the limit mechanism M^0 satisfies the exact incentive constraints needed to implement F. So M^0 implements F over T. □

Example 7. Here is an example of an SCC with one agent that is very strongly implementable over every (independent) non-EU extension, but not implementable by any ordinal mechanism.

Just consider the following minor modification of Example 1:

$t_1 : a \succ d \succ b \succ c$	$\{\frac{1}{2}a + \frac{1}{2}d\}$
$t_1' : b \succ c \succ a \succ d$	$\{\frac{1}{2}b + \frac{1}{2}c\}$
$t_1'' : a \succ b \succ c \succ d$	$\{\alpha(\frac{1}{2}a + \frac{1}{2}d) + (1-\alpha)(\frac{1}{2}b + \frac{1}{2}c) \mid \alpha \in [0,1]\}$

This SCC F is very strongly implementable over every non-EU extension of T_1 by the mechanism where agent 1 just picks his favorite among all the lotteries allowed by F. ◊

Example 8. Let X consist of 14 outcomes, $X = \{a, a', b, b', c, d, e, f, g, h, x, x', y, y'\}$. Building on Example 3, we take again three agents, with two ordinal types for agents 1 and 2, and three ordinal types for agent 3. The ordinal preferences are

$$t_1 : b \succ b' \succ f \succ a \succ a' \succ e \succ c \succ g \succ d \succ h \succ x \succ x' \succ y \succ y',$$
$$t_1' : x \succ x' \succ y \succ y' \succ h \succ d \succ g \succ c \succ e \succ b \succ b' \succ f \succ a \succ a',$$
$$t_2 : a \succ a' \succ c \succ b \succ b' \succ d \succ f \succ h \succ e \succ g \succ x \succ x' \succ y \succ y',$$
$$t_2' : x \succ x' \succ y \succ y' \succ g \succ e \succ h \succ f \succ d \succ a \succ a' \succ c \succ b \succ b',$$
$$t_3 : c, d, e, f, g, h \succ a \succ x \succ a' \succ x' \succ b \succ y \succ b' \succ y',$$
$$t_3' : c, d, e, f, g, h \succ b' \succ y' \succ a \succ x \succ a' \succ x' \succ b \succ y,$$
$$t_3'' : x \succ a \succ x' \succ a' \succ c, e, g \succ y' \succ b' \succ y \succ b \succ d, f, h.$$

(The preferences of agents 1 and 2 here are exactly as in Example 3, with x, x', y, and y' added in place of m.)

Let F be the simple SCC that specifies the following acceptable outcomes at each ordinal type profile:

t_3 :

	t_2	t_2'
t_1	a, b	c, d
t_1'	e, f	g, h

t_3':

	t_2	t_2'
t_1	a', b'	c, d
t_1'	e, f	g, h

t_3'':

	t_2	t_2'
t_1	x, x', y, y'	x, x', y, y'
t_1'	x, x', y, y'	x, x', y, y'

This F cannot be implemented by an ordinal mechanism. Such a mechanism would have to specify $\frac{1}{2} - \frac{1}{2}$ lotteries over the acceptable outcomes whenever agent 3 has ordinal type t_3 or t_3', and then there would be no way to choose the lottery at (t_1, t_2, t_3'') to satisfy the incentive constraints of all types of agent 3; the calculations are essentially identical to those in Example 3.

However, for any non-EU extension S of T, we can very strongly implement F over S as follows. If the agents report types such that 3's ordinal preferences correspond to t_3 or t_3', then carry out the $\frac{1}{2} - \frac{1}{2}$ lottery over the two outcomes allowed by F. Otherwise, ignore the types of agents 1 and 2, and carry out whichever of the lotteries

$$\alpha\left(\frac{1}{2}x + \frac{1}{2}y\right) + (1-\alpha)\left(\frac{1}{2}x' + \frac{1}{2}y'\right), \quad \alpha \in [0,1],$$

is most preferred by agent 3's non-EU type.

We now verify that this mechanism very strongly implements F.

If agent 3 has ordinal preferences t_3 or t_3', then every agent's ordinal incentive constraints are satisfied, so a fortiori the non-EU type has no incentive to deviate (including mixed-strategy deviations). If agent 3 has ordinal preferences t_3'', then the outcome is independent of the types reported by agents 1 and 2, so their incentive constraints are satisfied. It remains only to check the incentive constraint of each type s_3 of agent 3 with ordinal preferences t_3''. Suppose that s_3 is assigned its most preferred lottery

$$\pi(s_3) = \alpha\left(\frac{1}{2}x + \frac{1}{2}y\right) + (1-\alpha)\left(\frac{1}{2}x' + \frac{1}{2}y'\right).$$

Suppose agent 3 considers some (possibly mixed) deviation. Let $\sigma_{1,2} \in \Delta(T_1 \times T_2)$ denote 3's marginal belief about the others' ordinal types $(t_1(s_1), t_2(s_2))$, and let σ_3' be the marginal distribution over 3's ordinal types under the proposed deviation. This deviation would then lead to the lottery

$$\begin{aligned}
&\sigma_3'(t_3)\Big[\sigma_{1,2}(t_1,t_2)\left(\frac{1}{2}a + \frac{1}{2}b\right) + \sigma_{1,2}(t_1,t_2')\left(\frac{1}{2}c + \frac{1}{2}d\right) + \sigma_{1,2}(t_1',t_2)\left(\frac{1}{2}e + \frac{1}{2}f\right) \\
&\quad + \sigma_{1,2}(t_1',t_2')\left(\frac{1}{2}g + \frac{1}{2}h\right)\Big] \\
&\quad + \sigma_3'(t_3')\Big[\sigma_{1,2}(t_1,t_2)\left(\frac{1}{2}a' + \frac{1}{2}b'\right) + \sigma_{1,2}(t_1,t_2')\left(\frac{1}{2}c + \frac{1}{2}d\right) + \sigma_{1,2}(t_1',t_2)\left(\frac{1}{2}e + \frac{1}{2}f\right)
\end{aligned}$$

$$+\sigma_{1,2}(t_1',t_2')\left(\frac{1}{2}g+\frac{1}{2}h\right)\Bigg]$$
$$+\sigma_3'(t_3'')\left[\beta\left(\frac{1}{2}x+\frac{1}{2}y\right)+(1-\beta)\left(\frac{1}{2}x'+\frac{1}{2}y'\right)\right].$$

Here β is some number that represents an average of the lottery weights obtained by all the other non-EU types corresponding to t_3'' in the support of agent 3's random deviation.

This lottery is stochastically dominated for agent 3 by the lottery obtained from it by performing the replacements

$$a,c,e,g\mapsto x, \qquad a'\mapsto x',$$
$$b,d,f,h\mapsto y, \qquad b'\mapsto y'.$$

This latter lottery is equal to

$$\gamma\left(\frac{1}{2}x+\frac{1}{2}y\right)+(1-\gamma)\left(\frac{1}{2}x'+\frac{1}{2}y'\right)$$

with $\gamma=\sigma_3'(t_3)+\sigma_3'(t_3')(1-\sigma_{1,2}(t_1,t_2))+\sigma_3'(t_3'')\beta$. Finally, we know this is less preferred for s_3 than the originally assigned lottery $\pi(s_3)$, by construction of $\pi(s_3)$. Thus the proposed (random) deviation is not profitable for s_3, and the verification of very strong implementation is concluded. ♢

Proof of Lemma 1. Since the left-hand side of (9) is decreasing in q (for $q\le 1$), it suffices to hold fixed q' and show (9) at $q=q'-\epsilon$.

By moving the C^{k+1} term to the left side and multiplying out, (9) is equivalent to

$$\begin{aligned}&\left[2\delta\epsilon+\delta^2\left(1+q'^2-2q'\epsilon+\epsilon^2-2q'\right)\right]C^{k+1}-\left(1-\delta\left(q'-\epsilon\right)\right)^2\gamma\delta\\&\quad>\delta^2\left(1-q'\right)^2C^m+\left(2\delta q'-\delta^2q'^2\right)C^k.\end{aligned}\tag{18}$$

The left-hand side is bounded below by $(2\delta\epsilon-4\delta^2)C^{k+1}-4\gamma\delta$, and the right-hand side is bounded above by $\delta^2C^m+2\delta C^k$. Hence, multiplying (8) through by δ, we see that it implies (18) and, therefore, (9). □

Appendix B: Details on Bayesian implementation

We flesh out here some details on Bayesian implementation, originally sketched in Section 6.2.

For Bayesian implementation, type spaces should come equipped with priors. Suppose $T=T_1\times\cdots\times T_n$ is an ordinal type space, and suppose we are given, for each type t_i of each agent, a belief $p_{t_i}\in\Delta(T_{-i})$. (The family of beliefs $p=(p_{t_i})$ may come from a common prior over T, but this is not necessary.)

Given an SCC $F:T\rightrightarrows\Delta(X)$, we say that an ordinal mechanism $M:T\to\Delta(X)$ *Bayesian implements* F if $M(t)\in F(t)$ for each t, and for every agent i, all $t_i,t_i'\in T_i$, and

all outcomes $x \in X$,

$$\sum_{y \succeq_{t_i} x} \sum_{t_{-i} \in T_{-i}} p_{t_i}(t_{-i}) M(t_i, t_{-i})(y) \geq \sum_{y \succeq_{t_i} x} \sum_{t_{-i} \in T_{-i}} p_{t_i}(t_{-i}) M\big(t_i', t_{-i}\big)(y).$$

This latter condition is the criterion defined in previous literature as *ordinal Bayesian incentive compatibility*.

We next consider a cardinal extension S of T (without interdependence), $S = S_1 \times \cdots \times S_n$, together with priors $p_{s_i} \in \Delta(S_{-i})$ for each $s_i \in S_i$. For cardinal extensions in the Bayesian framework, we should impose a relationship between the cardinal priors and the ordinal priors: we say that the priors on S and T are *compatible* if, for each s_i, the distribution of $t_{-i}(s_{-i})$ induced by p_{s_i} equals the corresponding ordinal type's prior, $p_{t_i(s_i)}$.

Bayesian implementation on cardinal type spaces follows the usual definition: the mechanism $M : S \to \Delta(X)$ *Bayesian implements* F if $M(s) \in F(t(s))$ for all s, and for all i, s_i, s_i', we have

$$\sum_{s_{-i} \in S_{-i}} p_{s_i}(s_{-i}) s_i\big(M(s_i, s_{-i})\big) \geq \sum_{s_{-i} \in S_{-i}} p_{s_i}(s_{-i}) s_i\big(M\big(s_i', s_{-i}\big)\big).$$

As noted in the main text, if F is Bayesian implemented over T (with given priors) by the ordinal mechanism M, then for any cardinal extension S with compatible priors, F is Bayesian implemented over S by the mechanism $s \mapsto M(t(s))$.

The following example shows that the converse does not hold: F may be Bayesian implementable over every cardinal extension of T with compatible priors, but not implementable over T by an ordinal mechanism.

Consider two agents, with types t_1, t_1', and t_1'' for agent 1 and types t_2 and t_2' for agent 2. Let the outcome set be $X = \{a, b, c, d\}$ and let the types' preferences be

$$t_1 : a \succ d \succ b \succ c,$$
$$t_1' : c \succ a \succ d \succ b,$$
$$t_1'' : a \succ c \succ d \succ b,$$
$$t_2 : a \succ b \succ c \succ d,$$
$$t_2' : b \succ a \succ d \succ c.$$

Assume each ordinal type of each agent has a uniform belief over the ordinal types of the other agent, and let F be the simple SCC whose acceptable outcomes at each type profile are given by the table

	t_2	t_2'
t_1	a	b
t_1'	c	d
t_1''	a, b, c, d	a, b, c, d

For any cardinal extension of the type space T with compatible priors, F is Bayesian implementable—and by a deterministic mechanism at that. If agent 1's ordinal preferences are t_1 or t_1', then prescribe the outcome required by F. If agent 1's ordinal prefer-

ences are t_1'', then prescribe either a when 2's ordinal preferences are t_2 and b when t_2', or else c when t_2 and d when t_2', depending on which of the two lotteries $\frac{1}{2}a + \frac{1}{2}b$, $\frac{1}{2}c + \frac{1}{2}d$ is preferred by 1's cardinal type.

However, F is not Bayesian implementable by any mechanism M over the ordinal type space T: the incentive constraints of the various types of agent 1 cannot all be satisfied, by exactly the same argument as in Example 3.

We now move to allow interdependence in cardinal utilities. In this case, as stated in the main text, an SCC F may be Bayesian implementable over an ordinal type space T, yet not implementable over some interdependent extension S with compatible priors.

Here is a concrete example. Suppose there are two agents and two ordinal types of each agent, and the priors are uniform. Consider two outcomes $X = \{a, b\}$, and the following preferences and SCC (actually SCF) F:

	$t_2 : a \succ b$	$t_2' : b \succ a$
$t_1 : a \succ b$	a	b
$t_1' : b \succ a$	b	a

This obviously gives a unique ordinal mechanism, which is Bayesian incentive compatible (each agent gets the lottery $\frac{1}{2}a + \frac{1}{2}b$ no matter what type he reports).

However, consider an interdependent cardinal extension with corresponding types s_1, s_1', s_2, and s_2', and uniform priors. Suppose that the utility function of type s_1 is

$$u_1(a, s_1, s_2) = 1, \qquad u_1(b, s_1, s_2) = 0,$$
$$u_1\big(a, s_1, s_2'\big) = 2, \qquad u_1\big(b, s_1, s_2'\big) = 0.$$

Then for the unique mechanism consistent with F, type s_1 gets expected utility of $\frac{1}{2}$ from reporting s_1 but 1 from reporting s_1'. The mechanism is not incentive compatible. Thus F is not implementable.

Essentially, the problem is that once we let s_i's cardinal utility vary freely depending on s_{-i}, this freedom undoes any discipline imposed by the priors. (A version of this observation was previously made in Ledyard 1986.)

References

Abdellaoui, Mohammed (2002), "A genuine rank-dependent generalization of the von Neumann–Morgenstern expected utility theorem." *Econometrica*, 70, 717–736. [1297]

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2011), "Resolving conflicting preferences in school choice: The 'Boston' mechanism reconsidered." *American Economic Review*, 101, 399–410. [1276]

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2015), "Expanding 'choice' in school choice." *American Economic Journal: Microeconomics*, 7, 1–42. [1276]

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2005), "The New York City high school match." *American Economic Review Papers and Proceedings*, 95, 364–367. [1276]

Barberà, Salvador, Anna Bogomolnaia, and Hans van der Stel (1998), "Strategy-proof probabilistic rules for expected utility maximizers." *Mathematical Social Sciences*, 35, 89–103. [1303]

Bergemann, Dirk and Stephen Morris (2005), "Robust mechanism design." *Econometrica*, 73, 1771–1813. [1278, 1279, 1281, 1283, 1302]

Bhargava, Mohit, Dipjyoti Majumdar, and Arunava Sen (2015), "Incentive-compatible voting rules with positively correlated beliefs." *Theoretical Economics*, 10, 867–885. [1301]

Bodoh-Creed, Aaron L. (2012), "Ambiguous beliefs and mechanism design." *Games and Economic Behavior*, 75, 518–537. [1294]

Bogomolnaia, Anna and Hervé Moulin (2001), "A new solution to the random assignment problem." *Journal of Economic Theory*, 100, 295–328. [1275, 1276, 1277, 1292, 1303, 1304]

Bose, Subir, Emre Ozdenoren, and Andreas Pape (2006), "Optimal auctions with ambiguity." *Theoretical Economics*, 1, 411–438. [1294]

Carroll, Gabriel (2010), "An efficiency theorem for incompletely known preferences." *Journal of Economic Theory*, 145, 2463–2470. [1301]

Carroll, Gabriel (2012), "When are local incentive constraints sufficient?" *Econometrica*, 80, 661–686. [1301]

Chatterji, Shurojit, Arunava Sen, and Huaxia Zeng (2014), "Random dictatorship domains." *Games and Economic Behavior*, 86, 212–236. [1276, 1277, 1278, 1291, 1303]

Chatterji, Shurojit, Arunava Sen, and Huaxia Zeng (2016), "A characterization of single-peaked preferences via random social choice functions." *Theoretical Economics*, 11, 711–733. [1276, 1278]

Chatterji, Shurojit and Huaxia Zeng (2018), "On random social choice functions with the tops-only property." *Games and Economic Behavior*, 109, 413–435. [1276]

Chew, Soo Hong, Larry G. Epstein, and Uzi Segal (1991), "Mixture symmetry and quadratic utility." *Econometrica*, 59, 139–163. [1297]

Chung, Kim-Sau and Jeffrey C. Ely (2006), "Ex-post incentive compatible mechanism design." Unpublished paper, Northwestern University. [1278, 1282]

Chung, Kim-Sau and Jeffrey C. Ely (2007), "Foundations of dominant strategy mechanisms." *Review of Economic Studies*, 74, 447–476. [1283]

d'Aspremont, Claude and Bezalel Peleg (1988), "Ordinal Bayesian incentive compatible representations of committees." *Social Choice and Welfare*, 5, 261–279. [1301]

Dubra, Juan, Fabio Maccheroni, and Efe A. Ok (2004), "Expected utility theory without the completeness axiom." *Journal of Economic Theory*, 115, 118–133. [1301]

Ehlers, Lars (2002), "Probabilistic allocation rules and single-dipped preferences." *Social Choice and Welfare*, 19, 325–348. [1276]

Ehlers, Lars and Bettina Klaus (2003), "Probabilistic assignments of identical indivisible objects and uniform probabilistic rules." *Review of Economic Design*, 8, 249–268. [1276]

Ehlers, Lars, Dipjyoti Majumdar, Debasis Mishra, and Arunava Sen (2016), "Continuity and incentive compatibility in cardinal voting mechanisms." Unpublished paper, Cahier de recherche no 2016-04, Université de Montréal. Département de sciences économiques. [1279]

Ehlers, Lars, Hans Peters, and Ton Storcken (2002), "Strategy-proof probabilistic decision schemes for one-dimensional single-peaked preferences." *Journal of Economic Theory*, 105, 408–434. [1276]

Erdil, Aytek (2014), "Strategy-proof stochastic assignment." *Journal of Economic Theory*, 151, 146–162. [1275]

Filos-Ratsikas, Aris, Søren Kristoffer Stiil Frederiksen, and Jie Zhang (2014), "Social welfare in one-sided matchings: Random priority and beyond." In *SAGT 2014: Algorithmic Game Theory*, volume 8768 of Lecture Notes in Computer Science, 1–12, Springer, Berlin, Heidelberg. [1303]

Freixas, Xavier (1984), "A cardinal approach to straightforward probabilistic mechanisms." *Journal of Economic Theory*, 34, 227–251. [1303]

Gibbard, Allan (1977), "Manipulation of schemes that mix voting with chance." *Econometrica*, 45, 665–681. [1276]

Hylland, Aanund (1980), "Strategy proofness of voting procedures with lotteries as outcomes and infinite sets of strategies." Unpublished paper, The University of Oslo. [1303]

Jehiel, Philippe, Moritz Meyer-ter-Vehn, Benny Moldovanu, and William Zame (2006), "The limits of ex post implementation." *Econometrica*, 74, 585–610. [1278, 1279]

Katta, Akshay-Kumar and Jay Sethuraman (2006), "A solution to the random assignment problem on the full preference domain." *Journal of Economic Theory*, 131, 231–250. [1275]

Kesten, Onur (2010), "School choice with consent." *Quarterly Journal of Economics*, 125, 1297–1348. [1304]

Kőszegi, Botond and Matthew Rabin (2007), "Reference-dependent risk attitudes." *American Economic Review*, 97, 1047–1073. [1297, 1298]

Ledyard, John O. (1986), "The scope of the hypothesis of Bayesian equilibrium." *Journal of Economic Theory*, 39, 59–82. [1315]

Lopomo, Giuseppe, Luca Rigotti, and Chris Shannon (2014), "Uncertainty in mechanism design." Unpublished paper, University of Pittsburgh. [1294]

Machina, Mark (2008), "Non-expected utility theory." In *The New Palgrave Dictionary of Economics* (Steven N. Durlauf and Lawrence E. Blume, eds.), 74–84, Palgrave Macmillan, Basingstoke, United Kingdom. [1294]

Majumdar, Dipjyoti and Arunava Sen (2004), "Ordinally Bayesian incentive compatible voting rules." *Econometrica*, 72, 523–540. [1301]

Masatlioglu, Yusufcan and Collin Raymond (2016), "A behavioral analysis of stochastic reference dependence." *American Economic Review*, 106, 2760–2782. [1297]

Milgrom, Paul (2011), "Critical issues in the practice of market design." *Economic Inquiry*, 49, 311–320. [1279]

Mishra, Debasis (2016), "Ordinal Bayesian incentive compatibility in restricted domains." *Journal of Economic Theory*, 163, 925–954. [1301]

Pathak, Parag A. and Jay Sethuraman (2011), "Lotteries in student assignment: An equivalence result." *Theoretical Economics*, 6, 1–17. [1276]

Peters, Hans, Souvik Roy, Arunava Sen, and Ton Storcken (2014), "Probabilistic strategy-proof rules over single-peaked domains." *Journal of Mathematical Economics*, 52, 123–127. [1276]

Pycia, Marek (2014), "The cost of ordinality." Unpublished paper, SSRN 2460511, University of California, Los Angeles. [1276]

Pycia, Marek and M. Utku Ünver (2015), "Decomposing random mechanisms." *Journal of Mathematical Economics*, 61, 21–33. [1275]

Quiggin, John (1982), "A theory of anticipated utility." *Journal of Economic Behavior and Organization*, 3, 323–343. [1297]

Schummer, James (1999), "Strategy-proofness versus efficiency for small domains of preferences over public goods." *Economic Theory*, 13, 709–722. [1303]

Starmer, Chris (2000), "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk." *Journal of Economic Literature*, 38, 332–382. [1294]

Troyan, Peter (2012), "Comparing school choice mechanisms by interim and ex-ante welfare." *Games and Economic Behavior*, 75, 936–947. [1276]

Wakker, Peter P. (1994), "Separating marginal utility and probabilistic risk aversion." *Theory and Decision*, 36, 1–44. [1297]

Wolitzky, Alexander (2016), "Mechanism design with maxmin agents: Theory and an application to bilateral trade." *Theoretical Economics*, 11, 971–1004. [1294]

Yamashita, Takuro (2015), "Implementation in weakly undominated strategies: Optimality of second-price auction and posted-price mechanism." *Review of Economic Studies*, 82, 1223–1246. [1283]

Zhou, Lin (1990), "On a conjecture by Gale about one-sided matching problems." *Journal of Economic Theory*, 52, 123–135. [1303]

Co-editor Thomas Mariotti handled this manuscript.

Manuscript received 1 February, 2017; final version accepted 12 December, 2017; available online 13 December, 2017.