Abstract

A planner wants to extract information about an agent’s preference relation, but not necessarily the entire relation. Formally, a partition of the set of all possible orderings of alternatives (a ‘type space’) is given, and the planner wants to know to which partition element (‘type’) the agent’s true preference belongs. We say that a type space is elicitable if there exists a mechanism mapping types to (possibly random) outcomes in which the agent strictly prefers truth-telling over lying. In the Savage framework a type space is elicitable if and only if it can be elicited by offering the agent a list of menus and paying one randomly-chosen choice. When the planner can use objective lotteries, more type spaces can be elicited.

Keywords: Elicitation; incentive compatibility

JEL Classification: D8, C7

1 Introduction

In mechanism design and social choice settings, a planner collects information about agents’ preferences and uses this information to select a desired outcome. Often the entire preference relation is elicited, but in some settings the planner collects only partial information. For example, students in the New York City High School match are only asked to report their top 12 schools, rather than their complete ranking of hundreds of schools [1]. In that case information is truncated for simplicity. In other cases the planner simply doesn’t need the entire ordering to execute her objective. As a very simple example, suppose she wants to give a single agent his most-preferred alternative (meaning, she wants to implement the dictatorial social choice function). Here she only needs to ask the agent for his single most-preferred alternative, not his entire ranking of alternatives.

We refer to the information collected as the agent’s type. In the single-agent example, the agent’s type is simply which alternative he ranks at the top. More formally, if his most-preferred alternative

---

1Note that types here do not refer to hierarchies of beliefs as in the framework of Harsanyi. Rather, they correspond to the private information of the agent that the planner wants to elicit.

2We assume strict preferences.
is \( x \), then his type is the set of all preferences that rank \( x \) at the top. In the New York example, each of the student’s types contains all preference orderings that agree on the top-twelve options. The type is exactly what the planner elicits, nothing more and nothing less. And the set of all possible types forms a partition of the set of preferences: every preference ordering belongs to exactly one type. We therefore refer to this partition as the type space.

In this paper we explore what information about preferences—what type spaces—have strictly incentive compatible mechanisms. Formally, we ask for which partitions of preferences does there exist an incentive compatible method of eliciting the agent’s type. In the dictator example, paying the agent their announced favorite alternative is clearly incentive compatible. Thus, that particular type space is elicitable. Not all type spaces are elicitable, however. Consider three alternatives and the type space defined by the second-best element, meaning the planner wishes to know which of the three objects the agent ranks in the middle. We show that there is no incentive compatible payment mechanism on that type space. Thus, it is not an elicitable type space.

Our model assumes a finite number of alternatives and strict preferences. We allow the planner to use random mechanisms, which may pay an uncertain outcome. And we require that truthfully revealing one’s type be strictly optimal, regardless of risk preferences. Thus, a mechanism is strictly incentive compatible on a type space if, for every type, announcing their type truthfully pays a gamble that strictly dominates any gamble they could be paid by lying.

Our notion of elicitability requires that no extra information is learned about preferences beyond the agent’s type in the given type space. This is crucial, since the finest type space—in which every preference relation is its own type—is elicitable. To do so, the planner simply offers the agent every possible pair of alternatives, asks the agent to pick their favorite from each pair, and then pays them their chosen alternative from one randomly-selected pair. But if the type space is coarse, as in the dictatorial example above, this procedure elicits far more information than wanted by the planner. This may be undesirable because it violates the agent’s privacy by asking them to reveal more information than necessary. It is also inefficient, as the amount of information transferred (i.e., the number of questions the agent must answer) is much larger. Motivated by these considerations, we proceed assuming the planner wants only to know the type and nothing else.

We consider two ways of modelling the random payments generated by the mechanism: Lotteries or Savage-style acts. Lotteries represent the case of objective uncertainty, where the probabilities of the various outcomes are set by the planner. Acts, on the other hand, allow for subjective uncertainty. In the acts case the planner cannot control the payment probabilities. This limits their ability to construct incentive compatible mechanisms. Thus, there are type spaces that are elicitable with lotteries (using objective probabilities) but not with acts.

To illustrate, consider the type space based on the agent’s least-preferred alternative out of \( m \) possible alternatives. If the planner can use objective lotteries then this type space is elicitable. To do so, she asks the subject to announce his least-preferred alternative. If the agent reports \( x \), then the

\[^3\text{This is a description of an ‘indirect mechanism’, but clearly the same mapping from types to outcomes can be described as a direct mechanism: The agent reports his preference ordering over all alternatives, a pair of alternatives is randomly drawn, and the higher ranked of the two is awarded to the agent.}\]
planner gives him a lottery that pays every alternative except $x$ with probability $1/(m - 1)$, and pays $x$ with probability zero. Essentially, the planner asks the agent which one alternative he would like not to receive in the lottery. It is easy to see that announcing anything other than the true lowest-ranked alternative gives a (first-order) stochastically-dominated lottery to the agent. This is only true, however, because the probabilities on the $m - 1$ other alternatives are exactly equal. In the acts framework the planner would use a physical randomizing device (such as an $m - 1$-sided die) but could not be sure that the agent believes each outcome is equally likely. Incentive compatibility would not hold in this case since, regardless of the mapping from the state of the die to alternatives, the Savage act generated by truth-telling does not dominate state-by-state the acts generated by all possible lies.

To understand our characterization in the acts framework, consider the dictator example above. There the type is elicited by offering a single menu of all alternatives and paying the one chosen by the agent. We say that particular type space is generated by that single menu. In the case of eliciting the entire preference relation (the finest type space), the agent is offered all binary menus and one is randomly selected for payment. Thus, the finest type space is generated by the list of all binary menus. In general, we prove that a type space is elicitable if and only if it is generated by some list of menus. And the incentive compatible mechanism in that case is simple: offer exactly those menus to the agent and randomly select one for payment. The sufficiency of this condition is a simple extension of our previous work on incentives in experiments [2]; that this condition is also necessity is a new result.

In the lotteries framework we do not have a complete characterization for the general case. Clearly any partition generated by menus can also be elicited using lotteries, since lotteries are more powerful than acts. For necessity, we show that a convexity condition on partitions must be satisfied for a partition to be elicitable. Convexity is also sufficient in the important special case of type spaces that provide information only about the position of alternatives in the ranking.

All of our main analyses focus on the case of a single agent. In the final section we extend these to the case of multiple agents and focus on dominant strategy incentive compatibility and lottery payments. The extension is trivial: If each agent’s type space can be elicited individually, then their types can also be jointly elicited. To do so, simply run each individual’s incentive compatible mechanism and then take the equally-weighted convex combination of all lotteries generated. The resulting centroid lottery is paid to all agents. Each agent can only influence the resulting lottery to a small degree, but otherwise their incentives are exactly as in the individual case. Thus, regardless of what others announce, each will find truth-telling to strictly dominate any lie.

2 Notation and definitions

Let $X$ be a finite set of alternatives, with $|X| = m \geq 2$. Alternatives in $X$ will be denoted by $x, y, z, w$ etc. Let $O$ be the set of all strict orders (complete, transitive, antisymmetric binary relations) on $X$. Typical elements of $O$ are $\succeq, \succeq', \succeq''$, and typical subsets of $O$ are $A, A', A_1, A_2, \ldots$ etc. A collection of subsets $P = \{A_1, \ldots, A_k\}$ is a partition of $O$ if these sets are non-empty, pairwise disjoint, and
\[ \bigcup_{i=1}^{k} A_i = O. \] Given a partition \( P \) and \( \succeq \in O \), let \( P(\succeq) \) be the (unique) element of \( P \) that contains \( \succeq \).

### 2.1 Lotteries

Let \( \Delta(X) \) be the set of lotteries on \( X \). We use \( p, q, r \) etc. to denote elements of \( \Delta(X) \). If \( p \in \Delta(X) \) then \( p(x) \) is the probability assigned to the element \( x \in X \) by the lottery \( p \).

A lottery \( p \) First-Order Stochastically Dominates a lottery \( q \) relative to \( \succeq \) (denoted \( p \succeq^* q \)) if for every \( x \in X \)

\[
\sum_{\{y : y \succeq x\}} p(y) \geq \sum_{\{y : y \succeq x\}} q(y).
\]

If there is a strict inequality for at least one \( x \) then \( p \) strictly dominates \( q \) relative to \( \succeq \) (denoted \( p \succ^* q \)). It is worth pointing out the well-known fact that \( p \succeq^* q \) can be interpreted as the lottery \( p \) yielding (weakly) higher expected utility than the lottery \( q \) for every utility function \( u : X \rightarrow \mathbb{R} \) consistent with \( \succeq \) (i.e., for every \( u \) satisfying \( u[x] > u[y] \) whenever \( x \succeq y \) and \( x \neq y \)). This interpretation will be used extensively in Section 4 below when analyzing incentive compatibility.

Given a partition \( P \) of \( O \), a \( P \)-adaptable mechanism (in the lotteries framework) is a mapping \( g : P \rightarrow \Delta(X) \). The interpretation is that the agent announces an element of \( P \) and the mechanism outputs a lottery over \( X \). A mechanism is IC if the lottery obtained by announcing the element of \( P \) that contains the true preference dominates any other feasible lottery.\(^5\) Formally,

**Definition 1.** A \( P \)-adaptable mechanism \( g \) is IC if for every \( \succeq \in O \) and every \( A \neq P(\succeq) \)

\[ g(P(\succeq)) \succ^* g(A). \]

Our interest is in understanding which partitions \( P \) admit an IC \( P \)-adaptable mechanism, hence the following definition.

**Definition 2.** A partition \( P \) is elicitable (in the lotteries framework) if there exists an IC \( P \)-adaptable mechanism \( g : P \rightarrow \Delta(X) \).

### 2.2 Acts

In the acts framework we do not restrict the agent to view the randomization device of the mechanism as generating objective lotteries. Rather, the agent may have his own subjective beliefs about the likelihood of different realizations, or he may even have beliefs that cannot be represented by an additive probability measure. From the point of view of the agent the output of the mechanism is an act, i.e. a mapping from the state space of the randomization device to \( X \). By varying his announcement the agent controls which act he receives.

Formally, given a partition \( P \) of \( O \), a \( P \)-adaptable mechanism (in the acts framework) is a pair \((\Omega, f)\), where \( \Omega \) is a finite state space (of the randomization device) and \( f \) is a mapping \( f : P \rightarrow X^\Omega \).

\(^5\)Our notion of incentive compatibility is not completely standard as it requires strict rather than weak dominance. In particular, constant mechanisms are not incentive compatible according to our definition.
Definition 3. A $P$-adaptable mechanism $(\Omega, f)$ is IC if for every $\succeq \in O$ and every $A \neq P(\succeq)$

$$f(P(\succeq))(\omega) \succeq f(A)(\omega),$$

with a strict preference for some $\omega \in \Omega$.

Thus, incentive compatibility requires that the act obtained by truth-telling (reporting $P(\succeq)$ when the preference is $\succeq$) weakly dominates state-by-state any other feasible act, with a strict preference in at least one state. One way to interpret this is that every subjective expected utility agent with full-support beliefs strictly prefers truth-telling over lying. But Definition 3 guarantees that truthfulness is the unique best-response for other classes of preferences over acts as well (e.g., essentially any model in the Knightian uncertainty literature).

Definition 4. A partition $P$ is elicitable (in the acts framework) if there exists an IC $P$-adaptable mechanism $(\Omega, f)$.

3 Characterization in the acts framework

For every $\succeq \in O$ and a subset of alternatives $X' \subseteq X$, denote by $\text{dom}_\succeq(X')$ the maximal element in $X'$ according to $\succeq$. That is $\text{dom}_\succeq(X') \in X'$ and $\text{dom}_\succeq(X') \succeq x$ for every $x \in X'$.

Given a finite collection of subsets of alternatives ('menus') $X_1, \ldots, X_l \subseteq X$, say that $\succeq$ cannot be distinguished from $\succeq'$ based on $\{X_1, \ldots, X_l\}$ if $\text{dom}_\succeq(X_i) = \text{dom}_{\succeq'}(X_i)$ for every $i = 1, \ldots, l$. Clearly this defines an equivalence relation on $O$. Let $\tilde{P}(X_1, \ldots, X_l)$ be the partition of $O$ into the equivalence classes of this equivalence relation. Intuitively, $\tilde{P}(X_1, \ldots, X_l)$ contains the information revealed by observing the agent chooses his favorite element from each one of the menus $X_1, \ldots, X_l$.

Definition 5. A partition $P$ is generated by menus if there are $l$ and $X_1, \ldots, X_l \subseteq X$ such that $P = \tilde{P}(X_1, \ldots, X_l)$.

Proposition 1. A partition $P$ is elicitable in the acts framework if and only if it is generated by menus.

Proof. (If) Suppose $P = \tilde{P}(X_1, \ldots, X_l)$. Let $\Omega = \{\omega_1, \ldots, \omega_l\}$. For each $A \in P$ choose an arbitrary representative $\succeq^A \in A$, and define $f(A)(\omega_i) = \text{dom}_{\succeq^A}(X_i)$ for $i = 1, \ldots, l$. Note that by assumption the choice of the representative $\succeq^A$ does not affect the resulting mechanism.

To see that the above mechanism is IC fix some $\succeq$ and some $A \in P$. Then for each $i$ we have

$$f(P(\succeq))(\omega_i) = \text{dom}_{\succeq^P(\omega_i)}(X_i) = \text{dom}_{\succeq^A}(X_i) \succeq \text{dom}_{\succeq^A}(X_i) = f(A)(\omega_i),$$

where the first equality is by the definition of $f$, the second follows from $\succeq \in P(\succeq)$, the next relation follows from the definition of $\text{dom}$, and the last equality is again by construction of $f$. Moreover, if $A \neq P(\succeq)$ then there exists $i$ such that $\text{dom}_{\succeq^P}(X_i) \neq \text{dom}_{\succeq^A}(X_i)$ which gives a strict preference at $\omega_i$.

(Only If) Suppose $P$ is elicitable and let $(\Omega, f)$ be an IC $P$-adaptable mechanism. Enumerate the states so that $\Omega = \{\omega_1, \ldots, \omega_l\}$ for some positive integer $l$, and for each $i = 1, \ldots l$ define $X_i =$.
\{f(A)(\omega_i)\}_{A \in P} \subseteq X$. In words, $X_i$ is the set of all possible alternatives that can be chosen at state $\omega_i$ as the agent varies his announcement.

We now show that $P = \tilde{P}(X_1, \ldots, X_l)$. Suppose that $\succeq, \succ' \in \mathcal{P}$ are in the same element of $P$, and fix some $1 \leq i \leq l$. Then by incentive compatibility we have that $f(P(\succeq))(\omega_i) \geq f(A)(\omega_i)$ for every $A \in P$, which implies that $f(P(\succeq))(\omega_i) = \text{dom}_{\succeq}(X_i)$. Applying the same argument to $\succ'$ gives $f(P(\succ'))(\omega_i) = \text{dom}_{\succ'}(X_i)$. But since $P(\succeq) = P(\succ')$ we get $\text{dom}_{\succeq}(X_i) = \text{dom}_{\succ'}(X_i)$. Repeating for each $i = 1, \ldots, l$ shows that $\succeq, \succ'$ are in the same element of $\tilde{P}(X_1, \ldots, X_l)$.

Conversely, suppose that $\succeq, \succ' \in \mathcal{P}$ are in the same element of $\tilde{P}(X_1, \ldots, X_l)$. From the previous paragraph we have $f(P(\succeq))(\omega_i) = \text{dom}_{\succeq}(X_i)$ and $f(P(\succ'))(\omega_i) = \text{dom}_{\succ'}(X_i)$ for each $i$, so $f(P(\succeq))(\omega_i) = f(P(\succ'))(\omega_i)$ for each $i$. Incentive compatibility now implies that $P(\succeq) = P(\succ')$, which concludes the proof. 

\section*{3.1 Characterization of partitions generated by menus}

Since elicitable partitions in the acts framework are exactly those that are generated by menus, it is useful to have an ‘algorithm’ to determine whether a given partition is generated by menus. For a partition $P$ and a menu $X' \subseteq X$, say that $X'$ is identified by $P$ if for every $A \in P$ and every $\succeq, \succ' \in A$ it holds that $\text{dom}_{\succeq}(X') = \text{dom}_{\succ'}(X')$; in other words this means that $\text{dom}_{\succeq}(X')$ is a $P$-measurable function of $\succeq$ from $O$ to $X$. Let $\tilde{P}(P)$ be the collection of menus that are identified from $P$.

\begin{proposition}
A partition $P$ is generated by menus if and only if $P = \tilde{P}(\tilde{P}(P))$.
\end{proposition}

\begin{proof}
If $P = \tilde{P}(\tilde{P}(P))$ then clearly $P$ is generated by menus (by the menus $\tilde{P}(P)$).

To prove the converse note first that, for every partition $P$, if $\succeq, \succ' \in \mathcal{P}$ are in the same element $P$ then by definition $\text{dom}_{\succeq}(X') = \text{dom}_{\succ'}(X')$ for every $X' \in \tilde{P}(P)$. This implies that $\succeq, \succ'$ are also in the same element $\tilde{P}(\tilde{P}(P))$. In other words, $P$ is always (weakly) finer than $\tilde{P}(\tilde{P}(P))$.

Now, suppose that $P$ is generated by menus, say $P = \tilde{P}(X_1, \ldots, X_l)$. Then clearly $\{X_1, \ldots, X_l\} \subseteq \tilde{P}(P)$. But adding more menus can only make the resulting partition finer, so $\tilde{P}(\tilde{P}(P))$ is (weakly) finer than $\tilde{P}(X_1, \ldots, X_l) = P$. This completes the proof.
\end{proof}

\section*{4 Analysis in the lotteries framework}

\subsection*{4.1 A sufficient condition}

\begin{proposition}
If $P$ is generated by menus then $P$ is elicitable in the lotteries framework.
\end{proposition}

\begin{proof}
If $P$ is generated by menus then it follows from Proposition 1 that there is an IC $P$-adaptable mechanism $(\Omega, f)$ in the acts framework. Let $\mu$ be a full-support probability distribution on $\Omega$, and define the lotteries mechanism $g$ by

\[ g(A)(x) = \mu \left( \{ \omega \in \Omega : f(A)(\omega) = x \} \right) \]
for any $A \in P$ and $x \in X$. In words, $g(A)$ is the distribution of the $X$-valued random variable $f(A)$ when the state-space $\Omega$ is endowed with the measure $\mu$.

Now, fix $\succeq$ and $A \neq P(\succeq)$. Since $(\Omega, f)$ is IC we have that $f(P(\succeq))(\omega) \geq f(A)(\omega)$ for all $\omega$ and that $f(P(\succeq)) \neq f(A)$. Thus, for every $x \in X$

$$\{\omega \in \Omega : f(P(\succeq))(\omega) \geq x\} \supseteq \{\omega \in \Omega : f(A)(\omega) \geq x\},$$

with strict inclusion for at least one $x$. Since $\mu$ has full support it follows that $g(P(\succeq)) \succ g(A)$, and we are done. 

**Remark 1.** It is also possible to prove Proposition 3 directly. Suppose $P = \tilde{P}(X_1, \ldots, X_l)$. Let $\lambda$ be a full-support distribution on $\{1, \ldots, l\}$ and define $g(A)(x) = \lambda (\{1 \leq i \leq l : \text{dom}_\succeq\lambda(X_i) = x\})$, where $\succeq^A$ is an arbitrary choice from $A$. It is not hard to check that $g$ is IC.

The following example shows that, in contrast to the acts framework, $P$ may be elicitable even if it is not generated by menus. The intuition for this is that in the lotteries framework we make more restrictions than in the acts framework on the preferences that the agent may have over uncertain prospects, so incentive compatibility is easier to satisfy.

**Example 1.** Let $X = \{x, y, z\}$ and let $P = \{A_1, A_2, A_3\}$ where\(^6\)

$$A_1 = \{xyz, yxz\}, \quad A_2 = \{xzy, zxy\}, \quad A_3 = \{yzx, zyx\}.$$ 

In words, $P$ reveals the top 2 alternatives but not their order. Then $P$ is not generated by menus as can be easily checked using Proposition 2 above: $\tilde{I}(P)$ contains no non-trivial menus, so $\tilde{P}(\tilde{I}(P))$ is the coarsest partition and hence $\tilde{P}(\tilde{I}(P)) \neq P$. However, consider the mechanism $g$ given by

$$g(A_1) = (x, 0.5; y, 0.5; z, 0), \quad g(A_2) = (x, 0.5; y, 0; z, 0.5), \quad g(A_3) = (x, 0; y, 0.5; z, 0.5),$$

that is, $g$ chooses randomly one of the top-2 ranked alternatives. It is immediate to check that $g$ is IC.

**4.2 Necessary conditions**

Given a set of orderings $A \in O$, let $\succeq_A = \bigcap_{\succeq \in A} \succeq$ be the maximal relation that all orderings in $A$ agree on; that is, $x \succeq_A y$ if and only if $x \succeq y$ for all $\succeq \in A$. Note that $\succeq_A$ is a partial order. Say that $\succeq$ is consistent with $A$ if $\succeq_A \subseteq \succeq$, i.e., if for all $x, y \in X$, $x \succeq_A y$ implies $x \succeq y$. Denote by $\text{Cons}(A)$ the set of ordering consistent with $A$.

**Definition 6.** A set $A \subseteq O$ is convex if $A = \text{Cons}(A)$. A partition $P$ is convex if every $A \in P$ is convex.

To understand why we refer to the above property as convexity, recall that $u \in \mathbb{R}^X$ is consistent with an ordering $\succeq$ when $u[x] > u[y]$ iff $x \succeq y$ and $x \neq y$. Let $U(\succeq) \subseteq \mathbb{R}^X$ be the set of utility

\(^6\)The notation $xyz$ means the ordering that ranks $x$ first, $y$ second, and $z$ third. Other orderings are denoted in an analogous way.
vectors consistent with \( \succeq \). The closure \( \overline{U(\succeq)} \) of \( U(\succeq) \) is the set of utility vectors \( u \) satisfying \( u[x] \geq u[y] \) whenever \( x \succeq y \). Finally, for \( A \subseteq O \) we denote \( U(A) = \bigcup_{\succeq \in A} U(\succeq) \) and \( \overline{U(A)} = \bigcup_{\succeq \in A} \overline{U(\succeq)} \).

**Lemma 1.** A set \( A \subseteq O \) is convex if and only if \( \overline{U(A)} \) is convex in \( \mathbb{R}^X \).

**Proof.** It is not hard to check that for any set \( A \)

\[
\overline{U(\text{Cons}(A))} = \bigcap_{(x,y) : x \succeq y} \{ u : u[x] \geq u[y] \}.
\]

Since the set on the right-hand side is convex, it follows that if \( A \) is convex (i.e., \( A = \text{Cons}(A) \)) then \( \overline{U(A)} \) is convex. On the other hand, if \( U(A) \) is convex then it is an intersection of closed half-spaces of the form \( \{ u : u[x] \geq u[y] \} \). This implies that \( \overline{U(A)} = \overline{U(\text{Cons}(A))} \), and hence that \( A = \text{Cons}(A) \). \( \square \)

**Remark 2.** Convexity of a set of orderings can also be characterized using the notion of convexity of a set of nodes in a graph. For \( \succeq \in O \) and \( x \in X \) let \( r_\succeq(x) = |\{ y : y \succeq x \}| \) be the ranking of \( x \) in the ordering \( \succeq \). Say that two orderings \( \succeq \) and \( \succeq' \) are adjacent if there are \( x, y \in X \) such that \( r_\succeq(x) = r_{\succeq'}(x) = r_\succeq(y) - 1 = r_{\succeq'}(y) - 1 \) and \( r_\succeq(z) = r_{\succeq'}(z) \) for all \( z \neq x, y \), that is if \( \succeq \) and \( \succeq' \) differ only by a single switch of neighboring elements. Consider an undirected graph \( G \) where the set of vertices is \( O \) and the set of edges is the set of adjacent orderings. Then \( A \) is convex if and only if for every \( \succeq, \succeq' \in A \), if \( \succeq'' \) is in a shortest path between \( \succeq \) and \( \succeq' \) then \( \succeq'' \in A \). We omit the proof.

The following proposition shows that convexity is a necessary condition for elicatability in the lotteries framework. Variants of this result in different contexts have been obtained in previous works, see for example Lambert [5, pages 10-11] who attributes this observation to Osband [6].

**Proposition 4.** If \( P \) is elicitable in the lotteries framework then it is convex.

**Proof.** Let \( g \) be an IC \( P \)-adaptable mechanism, and let \( A \in P \). We will show that

\[
\overline{U(A)} = \bigcap_{A' \in P} \{ u : \langle g(A), u \rangle \geq \langle g(A'), u \rangle \},
\]

where \( \langle \cdot, \cdot \rangle \) is the standard scalar product in \( \mathbb{R}^X \). As the set on the right-hand side is clearly convex, this will suffice to prove the proposition.

Suppose first that \( u \) is in the set on the left-hand side. Then there is \( \succeq \in A \) such that \( u \in \overline{U(\succeq)} \). Incentive compatibility of \( g \) then implies that \( \langle g(A), u \rangle \geq \langle g(A'), u \rangle \) for every \( A' \in P \), so \( u \) is in the right-hand side as well.\(^7\)

Conversely, suppose that \( u \) is in the right-hand side. Then in every open neighborhood of \( u \) there is \( u' \) for which \( \langle g(A), u' \rangle > \langle g(A'), u' \rangle \) holds for all \( A' \neq A \) (here we use the fact that the right-hand side is a polyhedral set with non-empty interior). Incentive compatibility of \( g \) implies that \( u' \notin \overline{U(A')} \), so we must have \( u' \in U(A) \). Since this set is closed we get that \( u \in U(A) \) as well. \( \square \)

\(^7\)Recall that lottery \( p \) strictly dominates lottery \( q \) relative to \( \succeq \) if and only if \( \langle p, u \rangle > \langle q, u \rangle \) for every \( u \in U(\succeq) \).
Example 2. Let $X = \{x, y, z\}$ and let $P = \{A_1, A_2, A_3\}$ where
\[ A_1 = \{xyz, xzy, zxy\}, \quad A_2 = \{yxz, yzx\}, \quad A_3 = \{zyx\}. \]

Convexity of $P$ is immediate ($A_1$ is the set of all orderings in which $x$ is ranked above $y$, $A_2$ is the set of rankings in which $y$ is ranked above both $x$ and $z$, and $A_3$ is a singleton).

We show that $P$ is not elicitable. Suppose by contradiction that $g$ is an IC $P$-adaptable mechanism. Consider the vectors $u_1 \in U(A_1)$, $u_2 \in U(A_2)$ defined by $(u_1[x] = 2, u_1[y] = 1, u_1[z] = -M)$ and $(u_2[x] = 1, u_2[y] = 2, u_2[z] = -M)$, where $M > 0$ is large. Then incentive compatibility implies $\langle g(A_1) - g(A_2), u_1 \rangle > 0$ and $\langle g(A_1) - g(A_2), u_2 \rangle < 0$. Taking $M$ to $+\infty$ gives \( g(A_1)(z) = g(A_2)(z) \).

A similar argument (with $M$ replacing $-M$) shows that $g(A_1)(z) = g(A_3)(z)$. Combining these two equalities gives $g(A_2)(z) = g(A_3)(z)$.

Now, consider $u_3 \in U(A_3)$ given by $(u_3[x] = 1, u_3[y] = 2, u_3[z] = M)$ for large $M > 0$. From incentive compatibility $\langle g(A_2) - g(A_3), u_3 \rangle < 0$ and $\langle g(A_2) - g(A_3), u_2 \rangle > 0$, but this is inconsistent with $g(A_2)(z) = g(A_3)(z)$.

Roughly speaking, the reason that elicitation fails in Example 2 is that the cycle of sets $(A_2, A_1, A_3, A_2)$ is inconsistent with incentive compatibility: The probability of $z$ must be the same in $g(A_2)$ and $g(A_1)$, and must also be the same in $g(A_1)$ and $g(A_3)$, but at the same time must be higher in $g(A_3)$ than in $g(A_2)$. We now generalize this observation to obtain another necessary condition for elicitation.

Recall that $r_\geq(x) = |\{y : y \geq x\}|$ is the ranking of $x$ in the ordering $\geq$. Say that $\geq$ and $\geq'$ are adjacent via an $x - y$ switch if $r_\geq(x) = r_\geq(y) = r_\geq(y) - 1 = r_\geq(x) - 1$ and $r_\geq(z) = r_\geq(z)$ for all $z \neq x, y$. Similarly, say that the sets $A$ and $A'$ ($A \neq A'$) are adjacent via an $x - y$ switch if there are $\geq \in A$ and $\geq' \in A'$ that are adjacent via an $x - y$ switch; the sets $A$ and $A'$ are adjacent if they are adjacent via an $x - y$ switch for some $x, y$.

Proposition 5. Suppose that $P$ is elicitable. If $\{A_1, \ldots, A_k\} \subseteq P$ are such that $A_i$ is adjacent to $A_{i+1}$ for each $i \in \{1, \ldots, k\}$ (where addition is read modulo $k$), and if $A_1$ is adjacent to $A_2$ via an $x - y$ switch, then there exists $1 < i \leq k$ and $z$ such that $A_i$ and $A_{i+1}$ are adjacent via a $z - x$ switch.

Proof. The key to the proof is the following lemma.

Lemma 2. If $g$ is an IC $P$-adaptable mechanism, and if $A, A' \in P$ are adjacent via an $x - y$ switch, then $g(A)(x) - g(A)(y) = g(A')(y) - g(A')(x) > 0$, and $g(A)(z) = g(A')(z)$ for all $z \notin \{x, y\}$.

Proof. Let $\geq \in A$ and $\geq' \in A'$ be adjacent via an $x - y$ switch, that is $r_\geq(x) = r_\geq(y) = r_\geq'(y) - 1 = r_\geq'(x) - 1$ and $r_\geq(z) = r_\geq'(z)$ for all $z \neq x, y$.

We first show that $g(A)(z) = g(A')(z)$ for all $z$ with $r_\geq(z) < r_\geq(x)$, i.e., for all $z$ ranked above $x$ and $y$ (assuming such $z$ exists). The proof proceeds by induction on $r_\geq(z)$. For $r_\geq(z) = 1$, consider the utility vector $\bar{u}$ with $\bar{u}(z) = 1$ and $\bar{u}(w) = 0$ for all $w \neq z$. Then $\bar{u}$ is both a limit point of $U(\geq)$ and a limit point of $U(\geq')$. Any $u \in U(\geq)$ has $\langle u, g(A) \rangle > \langle u, g(A') \rangle$ and any $u' \in U(\geq')$ has $\langle u', g(A) \rangle < \langle u', g(A') \rangle$, from which we conclude that $\langle \bar{u}, g(A) \rangle = \langle \bar{u}, g(A') \rangle$ must be satisfied. But this is the same as $g(A)(z) = g(A')(z)$.
Now, consider $z$ with $r_{\geq}(z) < r_{\geq}(x)$ and suppose that $g(A)(w) = g(A')(w)$ for all $w$ for which $r_{\geq}(w) < r_{\geq}(z)$. Let $\bar{u}$ be given by $\bar{u}(w) = 1$ for all $w$ with $r_{\geq}(w) \leq r_{\geq}(z)$ and $\bar{u}(w) = 0$ otherwise. Observe again that $\bar{u}$ is both a limit point of $U(\succeq)$ and a limit point of $U(\succeq')$. Conclude that $\langle \bar{u}, g(A) \rangle = \langle \bar{u}, g(A') \rangle$, so by the induction hypothesis it follows that $g(A)(z) = g(A')(z)$.

A symmetric argument establishes the result when $r_{\geq}(z) > r_{\geq}(y)$ (e.g., for $r_{\geq}(z) = m$, use $\bar{u}(z) = 0$, $\bar{u}(w) = 1$ for $w \neq z$, and proceed by induction). Finally, since $g(A)(z) = g(A')(z)$ for all $z \neq x, y$, and since both $g(A), g(A')$ are lotteries, we must have $g(A)(x) - g(A)(y) = g(A')(y) - g(A')(x)$. The fact that these differences are positive immediately follows from incentive compatibility of $g$ (recall that $x$ is ranked above $y$ according to $\succeq$ and $y$ above $x$ according to $\succeq'$).

The proposition now easily follows. Indeed, let $g$ be an IC mechanism and suppose $\{A_1, \ldots, A_k\} \subseteq P$ satisfy the assumption of the proposition. Then by Lemma 2 we have that $g(A_1)(x) > g(A_2)(x)$. Suppose by means of contradiction that there is no $1 \leq i \leq k$ and $z$ such that $A_i$ and $A_{i+1}$ are adjacent via a $x \rightarrow y$ switch. Then it follows again from Lemma 2 that $g(A_i)(x) \geq g(A_{i+1})(x)$, whereby $g(A_2)(x) \geq g(A_1)(x)$, a contradiction.

Remark 3. Lemma 2 says that a necessary condition for a $P$-adaptable mechanism $g$ to be IC is that if $A$ and $A'$ are adjacent via an $x \rightarrow y$ switch, then the lotteries $g(A)$ and $g(A')$ are identical except that some mass is shifted from $x$ to $y$. This is a local incentive constraint which guarantees that an agent with true preference in $A$ has no incentive to announce $A'$, and vice versa. Carroll [3, Proposition 2] shows that in a class of models that includes ours, if a mechanism satisfies all the local incentive constraints then it is globally incentive compatible. He works with the standard notion of weak incentive compatibility, but the result goes through with our strict notion. Thus, the condition in Lemma 2 is not only necessary for $g$ to be IC, it is also sufficient.

Remark 4. An implication of Proposition 5 is that if $P$ is elicitable and $A, A' \in P$ are adjacent via an $x \rightarrow y$ switch, then these two sets can’t be adjacent via any other switch. Indeed, this corresponds to the case of a cycle of length 2 containing only $A$ and $A'$.

Remark 5. We do not know whether convexity together with the no-cycles condition of Proposition 5 is enough to guarantee elicitation in the lotteries framework. While we could not find a counter example, a problem may arise if several cycles of sets (each of which is not violating the condition) interact in a way that prevents a single mechanism to work for all of them simultaneously. Characterizing elicitable partitions in the lotteries framework is therefore still an open question.

4.3 Characterization for positional partitions

We now restrict attention to partitions that treat all the alternatives symmetrically, and only contain information about positions in the ranking. One example is when the agent only announces the ranking of his top $k$ alternatives, with $1 \leq k < m$. Another example is when the agent announces his $k$ lowest ranked alternatives (say, not including their order) to indicate what he views as unacceptable. We show below that in this class of partitions convexity is not only necessary but also sufficient for elicitation.
To formalize this, let us think of the ranking function \( r_\succ (\cdot) \) as a bijection from \( X \) to \( \{1, \ldots, m\} \). If \( B \subseteq \{1, \ldots, m\} \) then \( r_\succ^{-1}(B) = \{ x \in X : r_\succ(x) \in B \} \) is the set of alternatives whose ranking according to \( \succ \) is in \( B \). Given a partition \( Q \) of \( \{1, \ldots, m\} \), say that \( \succ, \succ' \) have the same \( Q \)-rankings if \( r_\succ^{-1}(B) = r_\succ'^{-1}(B) \) for every \( B \in Q \).

**Definition 7.** Let \( Q \) be a partition of \( \{1, \ldots, m\} \). The \( Q \)-positional partition, denoted \( P_Q \), is the partition of \( O \) into the equivalence classes of the equivalence relation ‘having the same \( Q \)-ranking’. A partition \( P \) is positional if it is \( Q \)-positional for some \( Q \).

**Example 3.** Suppose \( X = \{x, y, z\} \). For \( Q = \{\{1, 2\}, \{3\}\} \) the partition \( P_Q \) is the partition of Example 1, that is, \( P_Q = \{\{xyz, yxz\}, \{xzy, zyx\}, \{yxz, zyx\}\} \). For \( Q = \{\{1, 3\}, \{2\}\} \) we have \( P_Q = \{\{xyz, zyx\}, \{yxx, zyx\}, \{xzy, zyy\}\} \). For \( Q = \{\{1\}, \{2\}, \{3\}\} \) the partition \( P_Q \) is the finest partition (a unique ordering in each element). For \( Q = \{\{1, 2\}, \{3\}\} \) the partition \( P_Q \) is the coarsest partition, i.e., where all orderings are in a single element.

**Proposition 6.** Let \( P \) be a positional partition. Then the following conditions are equivalent:

1. \( P \) is elicitable in the lotteries framework.
2. \( P \) is convex.
3. Every element in the partition \( Q \) that defines \( P \) is a (possibly degenerate) interval in \( \{1, \ldots, m\} \).

**Proof.** (1) \( \implies \) (2): This follows from Proposition 4.

(2) \( \implies \) (3): We show that if (3) is violated then (2) is wrong as well. Suppose that \( 1 \leq i < j \leq k \leq m \) are such that \( i, k \in B \in Q \) but \( j \notin B \). Let \( \succ \) be an ordering such that \( r_\succ(x) = i, r_\succ(y) = j, \) and \( r_\succ(z) = k \) for some three elements \( x, y, z \in X \). Let \( \succ' \) be another ordering that is identical to \( \succ \) everywhere except that the rankings of \( x \) and \( z \) are switched; that is, \( r_\succ'(x) = k, r_\succ'(z) = i, \) and \( r_\succ'(w) = r_\succ(w) \) for all \( w \neq x, z \). Then by definition \( \succ \geq \succ' \) are in the same element of \( P_Q \), say \( A \).

Now, let \( u \in U(\succ) \) be such that \( u[x] = 3, u[y] = 1, \) and \( u[z] = 0 \). Existence of such \( u \) is obvious. Let \( u' \) be identical to \( u \) except that \( u'[x] = 0 \) and \( u'[z] = 3 \). Note that \( u' \in U(\succ') \). Consider \( u'' = \frac{1 + \epsilon}{3} u' + \frac{2 - \epsilon}{3} u \), where \( \epsilon > 0 \). We have \( u''[x] = 2 - \epsilon, \) \( u''[y] = 1 + \epsilon, \) and \( u'' \) is identical to \( u \) (and to \( u' \)) otherwise. Let \( \epsilon \) be sufficiently small such that no two elements of \( u'' \) are identical and such that no element of \( u'' \) is between 1 and \( 1 + \epsilon \). Call \( \succ'' \) to the ordering induced by \( u'' \). Then \( r_\succ''(x) = r_\succ(y) = j \), which implies that \( z \notin r_\succ''^{-1}(B) \). Thus, \( r_\succ''^{-1}(B) \neq r_\succ^{-1}(B) \), so \( \succ'' \notin A \). This proves that \( U(A) \) is not convex, so \( P_Q \) is not a convex partition.

(3) \( \implies \) (1): Suppose \( Q = \{B_1, \ldots, B_K\} \) where each \( B_k \) is an interval in \( \{1, \ldots, m\} \). Without loss assume that the \( B_k \)'s are ordered, so that if \( i \in B_k, j \in B_{k'} \) and \( k < k' \) then \( i < j \). Let \( \delta_1 > \delta_2 > \ldots > \delta_K \) be non-negative numbers satisfying \( \sum_{k=1}^K |B_k| \delta_k = 1 \). For each \( A \in P_Q \) choose an arbitrary ordering \( \succ^A \in A \). Define the mechanism \( g \) by \( g(A)(x) = \delta_k \) if \( r_\succ^A(x) \in B_k \). Note that by definition of the partition \( P_Q \) the mechanism \( g \) does not depend on the choice of the representatives \( \succ^A \). Also, \( g \) is IC since misreporting shifts probability from higher to lower ranked alternatives (recall Example 1).
5 Multiple agents

In this section we show that our analysis of elicitability can be extended straightforwardly to multi-agent setups. To make the point we focus here on the case of lotteries, but it should be clear that the results apply (with the necessary changes) to the acts framework.

Let \( N = \{1, \ldots, n\} \) be the set of agents. For each \( i \in N \) a partition \( P_i \) of \( O \) is given, and we let \( P = (P_1, \ldots, P_n) \) denote the profile of partitions. We use \( A_i \) for a typical element of \( P_i \), and \( A = (A_1, \ldots, A_n) \) for a profile of such elements. As usual, a subscript \( -i \) indicates that the \( i \)th coordinate of a vector is omitted.

A \( P \)-adaptable mechanism is a mapping \( g: P \rightarrow \Delta(X) \). Thus, for every \( A = (A_1, \ldots, A_n) \in P \) the lottery \( g(A) \in \Delta(X) \) is the output of the mechanism when each agent \( i \in N \) announces that his preference is in \( A_i \).

**Definition 8.** A \( P \)-adaptable mechanism \( g \) is Dominant-strategy IC (DIC) if for every \( i \in N \), every \( \succeq_i \in O \), every \( A_i \in P_i \) with \( A_i \neq P_i(\succeq_i) \), and every \( A_{-i} \in P_{-i} \)

\[
g(P_i(\succeq_i), A_{-i}) \succ_i g(A_i, A_{-i}).
\]

Notice that the above definition corresponds to the standard notion of a dominant-strategy mechanism, where truthfully reporting one’s type is optimal regardless of other agents’ reports. However, as in the previous sections, we require strict domination.

**Definition 9.** A profile of partitions \( P = (P_1, \ldots, P_n) \) is DIC-elicitable if there exists a DIC \( P \)-adaptable mechanism \( g: P \rightarrow \Delta(X) \).

**Proposition 7.** The profile of partitions \( P = (P_1, \ldots, P_n) \) is DIC-elicitable if and only if \( P_i \) is elicitable for each \( i \in N \).

**Proof.** If one of the \( P_i \)’s is not elicitable, then clearly \( P \) is not DIC-elicitable (just fix an arbitrary \( A_{-i} \)). In the other direction, suppose that every \( P_i \) is elicitable and let \( g_i: P_i \rightarrow \Delta(X) \) be a \( P_i \)-adaptable IC mechanism. For every \( A \in P \) define

\[
g(A) = \frac{1}{n} \sum_{i=1}^{n} g_i(A_i).
\]

Since \( g_i(P_i(\succeq_i)) \succ_i g_i(A_i) \), it follows that \( g(P_i(\succeq_i), A_{-i}) \succ_i g(A_i, A_{-i}) \) for any \( A_{-i} \), and the result follows.

**Remark 6.** A similar result to Proposition 7 holds if one replaces the notion of DIC by Bayes incentive compatibility. Namely, let \( \mu \) be a full-support product distribution over \( \times_{i \in N} P_i \). Given a \( P \)-adaptable mechanism \( g: P \rightarrow \Delta(X) \), \( i \in N \), and \( A_i \in P_i \), let \( E_{\mu_{-i}}[g(A_i, A_{-i})] \) be the expectation of \( g(A_i, A_{-i}) \) when \( A_{-i} \) is distributed according to the marginal of \( \mu \) on \( P_{-i} \). Say that \( g \) is Bayesian IC (BIC) if \( E_{\mu_{-i}}[g(P_i(\succeq_i), A_{-i})] \succ_i E_{\mu_{-i}}[g(A_i, A_{-i})] \) for every \( i \), every \( \succeq_i \), and every \( A_i \neq P_i(\succeq_i) \). Finally, say that \( P \) is BIC-elicitable under \( \mu \) if there exists a BIC \( P \)-adaptable mechanism \( g \). It is not hard to show that
$P$ is BIC-elicitable under $\mu$ if and only if each of the $P_i$’s is elicitable. We note that the assumption that $\mu$ is a product measure is important for this result.

References


