

When are Local Incentive Constraints Sufficient?

Gabriel Carroll, MIT

`gdc@mit.edu`

July 30, 2010

Abstract

In some mechanism design settings, it is plausible that agents cannot consider every possible misreport of their type. One can reasonably ask whether guarding only against some subset of likely misreports allows for more mechanisms than guarding against all possible misreports; if so, such a framework could provide a way to overturn existing negative results. Accordingly, we formulate a plausible notion of local incentive-compatibility in both continuous and discrete type spaces, in which agents only consider misreports that are close to their true type. We show that on many preference domains — including any convex domain of cardinal preferences, any domain of ordinal preference types with convex closure, single-peaked ordinal preferences, and successive single-crossing ordinal preferences — local incentive-compatibility actually implies full incentive-compatibility, for all probabilistic mechanisms. On domains of cardinal preferences that satisfy a strong nonconvexity condition, including the domain of single-peaked cardinal preferences, it does not. Our results provide a strengthening of many existing impossibility and characterization theorems.

Thanks to (in random order) Alex Wolitzky, Ron Lavi, Suehyun Kwon, Vince Conitzer, Anton Tsoy, Alex Frankel, Ariel Procaccia, Parag Pathak, Daron Acemoglu, Tim Roughgarden, and Rakesh Vohra for helpful comments and suggestions. The author is supported by an NSF Graduate Research Fellowship.

1 Introduction

Perhaps the most pervasive and fundamental concept in mechanism design is incentive-compatibility: no type of any agent should have an incentive to pretend to be any other type. This concept plays a crucial role in the characterization and impossibility results that comprise much of the axiomatic mechanism design literature. A typical such result shows that in some setting, the requirement of incentive-compatibility, combined with some other normative requirements, can be satisfied only by some specific class of mechanisms, or cannot be satisfied at all. This literature begins with the seminal Gibbard-Satterthwaite theorem [22, 43], showing that the only voting mechanisms with unrestricted ordinal preferences that are (dominant-strategy) incentive-compatible and can choose at least three different outcomes are dictatorships. Subsequent results include characterizations for voting with single-peaked preferences [34], single-crossing preferences [41], and multidimensional separable preferences [7]; impossibility results for two-sided stable matching [38, 2] and for efficient one-sided matching (assignment) [12, 46]; and many more.

However, in some settings, it is reasonable to think that an agent might not be able to consider all possible misreports of his preferences. For example, some mechanisms require stating a preference ordering over alternatives. In the Cambridge, Massachusetts, city council elections in 2009, there were 20 candidates (not counting write-ins), and a ballot consisted of a ranking of any subset of these candidates. In the recently redesigned New York City high school matching mechanism, each student may rank up to 12 preferred high school programs, out of over 500 available, and many students do list 12 programs [1]. In these cases, it is unlikely that an agent could contemplate all possible reorderings in order to check for advantageous misreports.

This raises the question of whether there are mechanisms in these settings that are not fully incentive-compatible, but do provide incentives against a few potentially salient misreports. If so, a mechanism designer might be willing to settle for this weaker form of incentive-compatibility, and existing characterization or impossibility results could conceivably be overturned.

In this paper, we give a general formulation of this kind of weakened incentive-compatibility, which we call *local incentive-compatibility*, the notion being that it guards against “small” misreports of preferences. After formulating our notion of mechanisms over completely general type spaces, we specialize separately to define local incentive-compatibility for mechanisms using cardinal preferences and for those using ordinal pref-

erences (actually we consider *polyhedral* type spaces, for which ordinal preferences are a special case). For cardinal preferences, we will say that a mechanism is locally incentive-compatible if no agent type can neither benefit from misreporting as some nearby type nor serve as a profitable misreport for some nearby type. For ordinal preferences, we will say that a mechanism is locally incentive-compatible if no agent type wants to misreport by switching some two consecutive outcomes in his preference ranking. Throughout most of the paper, we consider probabilistic mechanisms, which specify a distribution over outcomes as a function of the agent’s type. Randomness may be inherent in the mechanism, or it may arise subjectively for each agent via his uncertainty over other agents’ types. Of course, our results that hold for all probabilistic mechanisms also hold a fortiori for all deterministic mechanisms.

We show that for the type spaces encountered in many mechanism design problems — including any convex domain of cardinal preferences, domains of ordinal preferences with convex closure, single-peaked ordinal preferences, successive single-crossing ordinal preferences, and convex domains of preferences in settings with monetary transfers — local incentive-compatibility in our sense actually implies full incentive-compatibility, either for single-agent mechanisms or multi-agent mechanisms with independent types.

These results are in fact relevant for three reasons. The first is that they directly address the question already raised, which can be viewed as mechanism design in a boundedly rational framework: if a mechanism designer is concerned only about some prespecified set of likely misreports, rather than all conceivable misreports, does this enlarge the space of possible mechanisms? If the designer accepts our subset of misreports as being likely, then the answer is no. (Of course, if there are other misreports that are also considered likely, then the answer is still no; these results are robust to the possibility that our definition of local misreports is too stringent.) In particular, on these type spaces, requiring only local rather than full incentive-compatibility does not help in overcoming impossibility results.

The second reason is that there is a tradition in social choice theory of looking for the weakest assumptions necessary to obtain a characterization result. Our results immediately imply a weakening of the assumptions behind many results on incentive-compatible mechanisms, such as most of those mentioned in the first paragraph. Rather than assuming, say, dominant-strategy incentive-compatibility (together with some other normative criteria) to obtain a characterization result in some setting, we need only assume local incentive-compatibility, and the same conclusion ensues.

In particular, for ordinal type spaces, our results imply “computational” versions of many existing impossibility or characterization results. For example, the Gibbard-

Satterthwaite theorem says that any (deterministic) voting rule can be strategically manipulated at some preference profile, unless it is dictatorial or has at most two possible outcomes. The computational analogue states that any voting rule whose outcome is easy to compute can be easily manipulated at some preference profile, unless it is dictatorial or has at most two outcomes. Here by “easy” we mean computable in polynomial time as a function of the numbers of voters and outcomes, as in [8, 9]. The computational version of Gibbard-Satterthwaite follows from the original version and our result that local incentive-compatibility implies full incentive-compatibility: Any rule that is manipulable has some local manipulation, consisting of switching two candidates who are consecutive in the ranking, at some profile. So the manipulator can simply consider each such manipulation, starting from his true preferences, and compute the outcome of the voting rule; this computation can be done in polynomial time and will uncover an advantageous manipulation at least some of the time. Actually, stronger results extending this idea are already known (e.g. [29]; see also [17] for a similar result), but our results lead more generally to computational versions of many other existing characterization results by the same argument.

The third reason our results are relevant is that they can potentially provide a shortcut to proving incentive-compatibility for newly designed mechanisms in the future: on the appropriate type spaces, one only has to show that a mechanism is locally incentive-compatible rather than checking all possible misreports.

The aforementioned results show that, for many important type spaces, local incentive-compatibility implies full incentive-compatibility. On the other hand, there are type spaces where the implication does not hold. In particular, we show this for domains of cardinal preferences that satisfy a strong nonconvexity condition (essentially, that the intersection of the domain with some open half-space be disconnected), and we also give a separate proof and a more detailed investigation for dichotomous preferences [15]. These results would not be compelling if they heavily depended on the way we chose to define local misreports. However, our negative results in these domains are relatively strong, showing that mechanisms can be immune to a large fraction of misreports without being fully incentive-compatible. In particular, for the domain of dichotomous preferences, we show that almost all of the incentive-compatibility constraints are logically independent: for almost any incentive constraint, some mechanism violates it but satisfies all other constraints.

Our work connects with several existing papers on mechanism design under a subset of incentive constraints. Green and Laffont [25] consider a general setup in which the space

of messages that agents can send equals the space of types, with exogenous restrictions on which messages each type is capable of sending, and study when the revelation principle applies. Celik [16] and Sher and Vohra [44] consider specific mechanism design problems under subsets of incentive constraints; they work in settings with monetary transfers and focus on finding the revenue-maximizing mechanism. There does not appear to be prior literature asking the general question of when local and full incentive-compatibility coincide; although in independent work, Sato [42] has simultaneously discovered some of our results for deterministic mechanisms.

Our work also bears some formal connection to a spate of recent work in settings with monetary transfers and quasilinear utility. In such a setting, a rule mapping types to outcomes is *implementable* if there exists an accompanying payment function (mapping types to transfers) that makes truthful revelation incentive-compatible. There has recently been much interest in simple conditions ensuring that a rule is implementable, e.g. [40, 10, 4]. In particular, our work is similar in spirit to a paper by Archer and Kleinberg [3]. They show that local implementability (suitably defined) implies implementability, on any convex space of cardinal types. However, we focus mainly on mechanisms without monetary transfers. In Subsection 3.5, we do consider monetary transfers, but we show that local incentive-compatibility implies full incentive-compatibility for a given mechanism (consisting of an outcome rule and a payment function together), whereas Archer and Kleinberg consider whether a given outcome rule is implementable by *some* payment function. Thus, our result does not follow from the result of Archer and Kleinberg, nor vice versa.

Although our positive results make use of one particular definition of local incentive-compatibility, this paper does not advocate this definition as necessarily being most behaviorally relevant. Rather, the point is to explore one new avenue for thinking about the practically important question of whether there are interesting mechanisms that are difficult, but not impossible, to manipulate.

2 Framework

We begin by presenting our general framework and discussing interpretation and modeling issues. The concrete results will then occupy the following two sections.

2.1 Definitions

We will focus on incentives in a mechanism for an individual agent. Subsequently, we will show how the ideas extend straightforwardly to multi-agent mechanisms, as long as preferences are private (i.e. do not depend on other agents' types) and independently distributed across agents.

From the agent's point of view, a mechanism takes the agent's preferences as input and determines an outcome, or a probability distribution over outcomes. In a voting context, the outcomes might, say, be candidates; in a matching context, they might be jobs, schools, or roommates.

Different strands of the mechanism design literature have made different assumptions about the form of preferences the agent can announce. In contexts where monetary transfers are available, such as auctions or public projects, it is standard to assume agents announce their cardinal valuation for each of the possible outcomes. Cardinal preferences are also assumed in some of the voting and matching literature [5, 28, 46], but usually, these literatures assume that agents can only report an ordinal ranking of outcomes. In some contexts, agents are further restricted to rank a limited number of outcomes (such as the New York City school choice mechanism mentioned above). We will first give a unified conceptual treatment that covers all of these situations, then specialize to define local incentive-compatibility in specific settings.

The formal definitions are as follows. Let X , the *outcome space*, be any finite set; m will denote its cardinality. Let $\Delta(X)$ denote the space of lotteries over X . The agent is assumed to have expected utility preferences over lotteries. It will be convenient to think of both lotteries over X and utility functions as elements of the vector space $\mathbb{R}^m$; we will sometimes perform vector operations on them, such as addition or scalar multiplication. If the agent's utility function is u , his payoff from a lottery L is given by the inner product $u \cdot L$.

A *type* is a nonempty subset of $\mathbb{R}^m$. A *type space* is a set of pairwise disjoint types. We henceforth use the term *type space* in preference to *domain*: the latter term suggests only an exogenous restriction on the set of utility functions the agent may have, whereas our notion of a type space conveys both which utility functions are possible and which ones are mutually indistinguishable from a mechanism's point of view.

Given a type space T , a (*probabilistic*) *mechanism* is a function $f : T \rightarrow \Delta(X)$. Thus, the mechanism chooses a distribution over outcomes, based on the agent's (reported) type. The mechanism is *deterministic* if $f(t)$ is a degenerate lottery for each $t \in T$.

Given the outcome space and type space, an *incentive constraint* is an ordered pair of types. The interpretation of the constraint (t, t') is that a type t cannot benefit from misreporting as type t' . Accordingly, we say that a mechanism f *satisfies* an incentive constraint (t, t') if, for all $u \in t$, $u \cdot f(t) \geq u \cdot f(t')$; equivalently, $u \cdot (f(t) - f(t')) \geq 0$. A mechanism *satisfies a set* of incentive constraints if it satisfies every constraint in the set.

A mechanism that satisfies the full set of incentive constraints $T \times T$ is *fully incentive-compatible*. This is exactly the usual meaning of incentive-compatibility: no agent of any type can benefit from any misreport of his type.

A set S of incentive constraints is *sufficient* if every mechanism that satisfies the constraints S is fully incentive-compatible.

We highlight several important kinds of type spaces, and relevant sets of incentive constraints that plausibly capture the idea of small misreports.

- A type space T is *cardinal* if every type is a singleton. In this case, abusing notation, we will think of types as vectors and T as a subset of $\mathbb{R}^m$; thus, for example, we write $f(u)$ rather than $f(\{u\})$.

For a cardinal T , a set S of incentive constraints will be called *local incentive constraints* if, for every $u \in T$, there exists an open neighborhood N_u around u in T (with the relative topology) such that $(u, u') \in S$ and $(u', u) \in S$ for every $u' \in N_u$. A mechanism that satisfies some set of local incentive constraints will be called *locally incentive-compatible*. This definition is again intuitive: a mechanism is locally incentive-compatible if the agent never wants to misreport by a small amount. For example, if d is the Euclidean distance function on $\mathbb{R}^m$ (or any other distance function that generates the same topology) and ϵ is a small positive number, then $\{(u, u') \mid d(u, u') < \epsilon\}$ is a set of local incentive constraints.

An alternative, perhaps more conventional, definition of local incentive constraints would simply be that for each $u \in T$, there should exist an open neighborhood N_u such that $(u, u') \in S$ for all $u' \in N_u$. That is, each type u has no incentive to make nearby misreports, but no relationship is assumed between the meanings of “nearby” for different types. In Subsection 3.1, we will indicate how this definition would give different results.

- A type space is *ordinal* if every type is of the form $t = \{u \mid u(x_1) > u(x_2) > \cdots > u(x_m)\}$ for some strict ordering $x_1 \succ \cdots \succ x_m$ of the elements of X . We say that t *represents* this ordering. When types are ordinal, f satisfies a constraint (t, t') if

and only if the lottery $f(t)$ first-order stochastically dominates $f(t')$ with respect to the ordering on X represented by t . (This is easy to show.)

Call two ordinal types t, t' *adjacent* if the orderings they represent differ only by a switch of two outcomes that are consecutive in the orderings represented by t and t' . For example, suppose the outcome set is $\{x, y, z\}$, and we write, say, t_{xyz} to denote the type representing the ordering $x \succ y \succ z$. Then t_{xzy} is adjacent to t_{zxy} and t_{xyz} but not to t_{yzx} . On an ordinal type space T , the *local incentive constraints* will refer to the set of all constraints (t, t') such that t and t' are adjacent. We will again say that a mechanism is *locally incentive-compatible* if it satisfies the local incentive constraints.

- More generally, we can consider *polyhedral* type spaces. In the space of utility functions, $\mathbb{R}^m$, an *open half-space* is a set of the form $\{u \mid u \cdot \lambda > c\}$ for some nonzero $\lambda \in \mathbb{R}^m$ and some constant c . If H is such an open half-space, its closure $\text{cl}(H) = \{u \mid u \cdot \lambda \geq c\}$ is a *closed half-space*, and its boundary $\partial H = \{u \mid u \cdot \lambda = c\}$ is a *hyperplane*. Define an (open) *polyhedron* to be a nonempty set that is the intersection of finitely many open half-spaces. A *polyhedral type space* is a type space consisting of finitely many types that are all polyhedra.

Define two disjoint polyhedra t, t' to be *adjacent* if $\text{cl}(t) \cap \text{cl}(t')$ contains a nonempty, relatively open subset of a hyperplane. (A set is relatively open in a hyperplane if it is open in the relative topology of the hyperplane.) Intuitively, t and t' are polyhedra that border along a face.

When the type space T is polyhedral, the *local incentive constraints* are the constraints (t, t') such that t and t' are adjacent. Again, a mechanism is *locally incentive-compatible* if it satisfies the local incentive constraints.

Any ordinal type space is polyhedral (the type $\{u \mid u(x_1) > \dots > u(x_m)\}$ is the intersection of the open half-spaces $\{u \mid u(x_i) - u(x_{i+1}) > 0\}$ for $i = 1, \dots, m-1$). In this case, one can check that the definitions of adjacency and local incentive constraints for ordinal types agree with those for polyhedral types. For another example, we have the types given by truncated rankings, i.e. $\{u \mid u(x_1) > \dots > u(x_k) \text{ and } u(x_k) > u(y) \text{ for all } y \neq x_1, \dots, x_k\}$, for any distinct outcomes $x_1, \dots, x_k$ (where $k < m$ is fixed). These are again polyhedral types; the type corresponding to a sequence $x_1, \dots, x_k$ is adjacent to all types that can be obtained by either switching x_i and x_{i+1} for some i , or replacing x_k with some $y \neq x_1, \dots, x_k$.

General polyhedral type spaces come up in the work of Gibbard [24], who considers probabilistic voting schemes with finite strategy sets, such that every voter is guaranteed to have a dominant strategy no matter what his cardinal utility function is. For each strategy, the set of utility functions for which it is dominant forms a polyhedron (ignoring boundary issues).

We are interested in identifying sufficient sets of incentive constraints, and in particular in determining whether or not local incentive constraints are sufficient, on various type spaces.

As mentioned earlier, while all of our analysis is from the point of view of a single agent, our results apply to multi-agent mechanisms. That is, if a set of incentive constraints is sufficient for one-agent mechanisms, it is also sufficient for multi-agent mechanisms, as can be seen by holding fixed the types of all agents except the one who is considering a misreport.

To make this more precise, define a mechanism with n agents, type space T , and outcome space X to be a map $f : T^n \rightarrow (\Delta(X))^n$, specifying a (probabilistic) outcome for each agent as a function of all the agents' types. (In specific settings the space of outcome n -tuples might be more restricted: for example, in a voting mechanism, all agents must get the same outcome.) Let $f_i : T^n \rightarrow \Delta(X)$ denote the i th component of f . Suppose that a set S of incentive constraints is sufficient for T . Now suppose f satisfies S for each agent i and each profile t_{-i} of types of the other agents: if $(t_i, t'_i) \in S$, then for each $u_i \in t_i$, $u_i \cdot (f_i(t_i, t_{-i}) - f_i(t'_i, t_{-i})) \geq 0$. Holding fixed the profile t_{-i} of other agents' types, then, the mechanism $t \mapsto f_i(t, t_{-i})$ satisfies the constraints S ; by sufficiency it is fully incentive-compatible. So $u_i \cdot (f_i(t_i, t_{-i}) - f_i(t'_i, t_{-i})) \geq 0$ for all $t_i, t'_i \in T$ and $u_i \in t_i$: the mechanism is fully incentive-compatible (in dominant strategies).

It is in this sense that S is sufficient to ensure full incentive-compatibility in multi-agent mechanisms. And the same argument applies for other notions of incentive-compatibility, such as Bayes-Nash, rather than dominant-strategy incentive-compatibility, as long as the agents' types are independently distributed. (Without independence, the argument fails, since we cannot vary t_i freely while holding fixed the distribution of t_{-i} .)

2.2 Indirect mechanisms

The framework given above focuses on direct mechanisms; in particular, incentive constraints state directly which types are to be prevented from imitating which other types. One could instead consider indirect mechanisms, in which each type of agent announces

a message that is then mapped to a distribution over outcomes. This approach could potentially give a mechanism designer much more freedom. As a bounded-rationality assumption, one might imagine that agents consider only misreports that are local in the *message space*; thus, incentive constraints consist of pairs of messages, and the definition of locality in a given message space is independent of the mechanism chosen. In this subsection we outline how a modeling framework along these lines might look and why it does not seem to be a promising avenue of study.

Given an outcome space X and type space T , an *indirect mechanism* consists of a *message space* M ; a *strategy* $\sigma : T \rightarrow M$; and an *outcome function* $g : M \rightarrow \Delta(X)$. An *incentive constraint* is now an ordered pair of elements of M , and a mechanism (M, σ, g) *satisfies* the incentive constraint (m, m') if, for all $t \in T$ such that $\sigma(t) = m$ and all $u \in t$, we have $u \cdot (g(m) - g(m')) \geq 0$. (If there are no t such that $\sigma(t) = m$, then the constraint is vacuously satisfied.)

The definition of local incentive constraints would depend on the structure of the space M . One natural approach, mirroring our earlier definition for cardinal type spaces, would be to require that M be a subset of some Euclidean space; then a set S of incentive constraints consists of *local incentive constraints* if every $m \in M$ has a neighborhood N_m such that $(m, m') \in S$ and $(m', m) \in S$ for all $m' \in N_m$.

We would then say that a function $f : T \rightarrow \Delta(X)$ is *locally incentive-compatibly implementable* if there exists some mechanism (M, σ, g) satisfying a set of local incentive constraints, such that $g \circ \sigma = f$. The function is *fully incentive-compatibly implementable* if some such mechanism satisfies the full set of incentive constraints $M \times M$; by the revelation principle, this is equivalent to full incentive-compatibility of the direct mechanism f . The question would then be: when are there functions f that are locally, but not fully, incentive-compatibly implementable?

Unfortunately, this approach gives trivial results. For example, if the image of f is finite (as when T is a polyhedral type space, or whenever f is a deterministic mechanism), then f is necessarily locally incentive-compatibly implementable. Indeed, let M be a discrete subset of Euclidean space, with enough elements so that there exist $\sigma : T \rightarrow M$ and $g : M \rightarrow \Delta(X)$ with $g \circ \sigma = f$. Then the set $\{(m, m) \mid m \in M\}$ is a set of local incentive constraints, which is always satisfied. Even if, say, we require M to be convex and open, any f is locally incentive-compatibly implementable: choose σ with finite image and $g : \text{Im}(\sigma) \rightarrow \Delta(X)$ so that $g \circ \sigma = f$, then extend g to M so as to be locally constant near each point of the image of σ . This mechanism satisfies any local incentive constraints as long as the neighborhood N_m of any $m \in \text{Im}(\sigma)$ is small enough that g is constant

there. The prediction that so many functions are implementable indicates that this model is unreasonable.

We could try to tweak the framework to avoid such excessive implementability results. For example, if T is a cardinal type space, we could restrict attention to mechanisms for which the strategy σ is continuous. But in this case, whenever an indirect mechanism (M, σ, g) satisfies the local incentive constraints S , the corresponding direct mechanism $g \circ \sigma$ satisfies the local incentive constraints $\{(t, t') \mid (\sigma(t), \sigma(t')) \in S\}$. So allowing indirect mechanisms would not buy us any new local incentive-compatibility that we could not already get with direct mechanisms.

The other problem with formulating local incentive constraints on the message space, rather than directly on the type space, lies with the implicit behavioral assumption that agents' choice of which misreports to consider does not depend on the choice of σ . It is not clear why such an assumption would be valid. For example, consider some direct mechanism f on the full ordinal type space T . Suppose that f was not fully incentive-compatible, but a designer tried to prevent computationally limited agents from manipulating it by using an indirect mechanism (M, σ, g) , with $g \circ \sigma = f$. An agent could then look for manipulations as follows: if the agent's true type is t , then for each of the $m - 1$ types t' adjacent to t in T , calculate $g(\sigma(t'))$ and compare it to $g(\sigma(t))$. This calculation would succeed in finding manipulations in at least some instances — even if the designer had taken pains to make the mechanism satisfy a large set of incentive constraints in M .

For these reasons, we continue to restrict attention to direct mechanisms, with local incentive constraints formulated on the type space T .

3 Sufficiency results

In this section we shall show that local incentive constraints are sufficient — any locally incentive-compatible mechanism is fully incentive-compatible — on a variety of common type spaces.

3.1 Cardinal type spaces

As already mentioned, in a cardinal type space, we can identify each type with the single utility function it contains, and thus identify the whole type space with a subset of $\mathbb{R}^m$, the Euclidean space of possible utility functions. This allows us to formulate our first sufficiency result:

Proposition 1 *On a convex cardinal type space T , any set of local incentive constraints is sufficient.*

The proof of this proposition is a model for the proofs of most of our other sufficiency results. To prove that an agent of type u never wants to misreport as type v , we restrict attention to types along the line segment $[u, v]$ connecting u and v , effectively collapsing the type space to one dimension; we then apply the kind of supermodularity or “revealed-preference” argument that is familiar elsewhere in the mechanism design literature (see e.g. [35, Lemma 2], [32, pp. 177-178], [37, Theorem 1]). Suppose u_α and u_β are two types along this line segment, with u_α closer to u than u_β , such that neither has an incentive to imitate the other under mechanism f . Since u_α has a stronger relative preference than u_β does for outcomes preferred by u , it must be that $f(u_\alpha)$ is preferred by u over $f(u_\beta)$. Breaking up $[u, v]$ into such shorter segments $[u_\alpha, u_\beta]$ and combining the resulting inequalities, we conclude that u prefers $f(u)$ over $f(v)$.

Proof: Let S be a set of local incentive constraints and f a mechanism satisfying S . For types u, v , write $u \leftrightarrow v$ if (u, v) and (v, u) are both in S . By definition, for every $u \in T$, there is some neighborhood N_u of u in T such that $u \leftrightarrow v$ for all $v \in N_u$.

Fix arbitrary $u, v \in T$. We want to show that $u \cdot (f(u) - f(v)) \geq 0$.

For any $\alpha \in [0, 1]$, define $u_\alpha = (1 - \alpha)u + \alpha v$. Convexity implies $u_\alpha \in T$ for each α . Let

$$A = \{\alpha \mid \text{there exist } 0 = \alpha_0 < \alpha_1 < \dots < \alpha_k \leq 1 \text{ with } u_{\alpha_0} \leftrightarrow u_{\alpha_1} \leftrightarrow \dots \leftrightarrow u_{\alpha_k} \text{ and } \alpha_k = \alpha\}.$$

Clearly, if $\alpha \in A$, $\alpha < \alpha' \leq 1$, and $u_\alpha \leftrightarrow u_{\alpha'}$, then $\alpha' \in A$. Now let $\bar{\alpha} = \sup A$ (we know A is nonempty since $0 \in A$). If $\bar{\alpha} > 0$, then for α sufficiently close to $\bar{\alpha}$ we have $u_\alpha \leftrightarrow u_{\bar{\alpha}}$; since we can choose $\alpha \in A$ arbitrarily close and less than $\bar{\alpha}$, we get $\bar{\alpha} \in A$. Moreover, if $\bar{\alpha} < 1$, then $u_{\bar{\alpha}} \leftrightarrow u_\alpha$ for α just slightly larger than $\bar{\alpha}$; this implies $\alpha \in A$, which contradicts $\bar{\alpha} = \sup A$. Therefore, we get $\bar{\alpha} = 1$ and $1 \in A$.

So we have $0 = \alpha_0 < \alpha_1 < \dots < \alpha_k = 1$ with $u_{\alpha_i} \leftrightarrow u_{\alpha_{i+1}}$ for each i . Now write out the local incentive constraints:

$$\begin{aligned} u_{\alpha_i} \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) &\geq 0, \\ u_{\alpha_{i+1}} \cdot (f(u_{\alpha_{i+1}}) - f(u_{\alpha_i})) &\geq 0. \end{aligned}$$

Multiplying by α_{i+1} and α_i , respectively, and adding gives

$$[\alpha_{i+1}u_{\alpha_i} - \alpha_i u_{\alpha_{i+1}}] \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) \geq 0.$$

But simple calculation shows that

$$\alpha_{i+1}u_{\alpha_i} - \alpha_i u_{\alpha_{i+1}} = (\alpha_{i+1} - \alpha_i)u,$$

hence

$$[(\alpha_{i+1} - \alpha_i)u] \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) \geq 0.$$

Since $\alpha_{i+1} - \alpha_i > 0$, we can divide through and actually get

$$u \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) \geq 0.$$

Now we can sum over $i = 0, 1, \dots, k-1$, and telescoping gives

$$u \cdot (f(u) - f(v)) = u \cdot (f(u_{\alpha_0}) - f(u_{\alpha_k})) \geq 0.$$

□

Proposition 1 applies to any convex cardinal type space. This includes, for example, the full space of utility functions (on any finite X); or the space of utility functions that are increasing with respect to some partial order on X ; or the space of supermodular utility functions, given a lattice structure on X ; or the space of utility functions satisfying some concavity conditions.

The proof of Proposition 1 clearly uses both parts of the definition of local incentive constraints — that for all u , there should exist a neighborhood N_u with both $(u, u') \in S$ and $(u', u) \in S$ for $u' \in N_u$. A weaker requirement, that only (u, u') need be in S , would be more intuitive, but this requirement is not enough. For example, suppose $X = \{x, y\}$ and T is the full space of all cardinal types. Consider the mechanism f given by $f(u) = x$ if $u(x) < u(y)$ and $f(u) = y$ otherwise. This f meets the weaker definition of local incentive-compatibility, but is not fully incentive-compatible.

An alternative way of expressing our definition of local incentive constraints is as follows: f is locally incentive-compatible if every $u \in T$ is a local maximum of both the functions $v \mapsto u \cdot f(v)$ and $v \mapsto v \cdot (f(u) - f(v))$. With this interpretation, local incentive-compatibility can potentially be checked by first- and second-order conditions when f is differentiable.

We can strengthen Proposition 1 slightly to allow the type space to exclude utility functions that are indifferent among all outcomes. Formally, say that a type u is *indifferent* if $u(x) = u(y)$ for all $x, y \in X$. Also, say that two types u, u' are *equivalent* if $u = \alpha u' + v$ for some scalar $\alpha > 0$ and some indifferent v (this is our usual notion of cardinal equivalence).

Corollary 2 *Suppose that T is a cardinal type space such that $T \cup T'$ is convex, for some set T' consisting of indifferent types. Suppose that T contains at least three mutually non-equivalent types. Then any set of local incentive constraints is sufficient.*

Proof: Let S be a set of local incentive constraints and f a mechanism satisfying them. Fix $u, v \in T$. We want to show that $u \cdot (f(u) - f(v)) \geq 0$.

If the line segment from u to v does not contain any indifferent types in its interior, then it must lie entirely in T , and the argument from the proof of Proposition 1 carries over unchanged. So suppose the segment does pass through some indifferent type w ; then we have $v = (1 + \alpha)w - \alpha u$ for some $\alpha > 0$. In particular, v is equivalent to $-u$ (and u is equivalent to $-v$).

By assumption, there exists some type $u' \in T$ that is not equivalent to u or v . Hence, the segment from u to u' does not contain any indifferent types in its interior, and so lies entirely in T . Likewise, so does the segment from v to u' . Thus, by the argument for Proposition 1, we have

$$u \cdot (f(u) - f(u')) \geq 0; \quad v \cdot (f(v) - f(u')) \geq 0.$$

However, v is equivalent to $-u$, so the latter of these readily implies $-u \cdot (f(v) - f(u')) \geq 0$. Hence,

$$u \cdot (f(u) - f(v)) = [u \cdot (f(u) - f(u'))] + [-u \cdot (f(v) - f(u'))] \geq 0.$$

□

The requirement of at least three mutually non-equivalent types cannot be removed. For example, suppose again that $X = \{x, y\}$, and consider the cardinal type space $T = \{u \mid u(x) \neq u(y)\}$, which has only two non-equivalent types. The mechanism f that always chooses the less-preferred outcome is locally, but not fully, incentive-compatible.

3.2 Polyhedral type spaces

Next we consider polyhedral type spaces, which includes ordinal type spaces as a special case. Our main result here is:

Proposition 3 *Let T be a polyhedral type space such that $\cup_{t \in T} \text{cl}(t)$ is convex. Then the set of local incentive constraints is sufficient.*

The argument is essentially the same as for Proposition 1: for utility functions u and v , we consider the line segment connecting u and v ; this segment passes through various types in succession. By jiggling v a bit if necessary, we can ensure that any two successive types along this line segment are adjacent polyhedra, and then we can just add up the corresponding local incentive constraints as before.

This is illustrated in Figure 1 for the case where T is the full ordinal type space on three outcomes. In this case, the types consist of six triangular-prism-shaped sectors of $\mathbb{R}^3$; by projecting orthogonally onto the plane $u(x_1) + u(x_2) + u(x_3)$ we get six sectors of the plane, as illustrated. Each type consists of the interior of one sector. The figure illustrates the proof that $u \cdot (f(t) - f(t')) \geq 0$, for a particular choice of t, t' and $u \in t$. We just choose $v \in t'$ and apply the argument from Proposition 1. If v happens to be inauspiciously chosen, such as the v' shown, the argument fails because the segment from u to v passes through two nonadjacent types in succession; but in this case we can just perturb v to get rid of the problem.

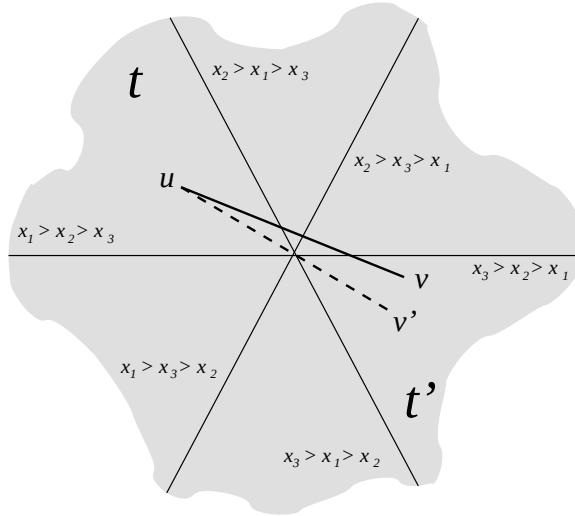


Figure 1: Argument for Proposition 3

Expressing this argument precisely requires a little technical work. For the remainder of this subsection, we fix the polyhedral type space T , and let S be the corresponding set of local incentive constraints.

Lemma 4 *Let the polyhedron t be written as an intersection of open half-spaces: $t =$*

$\cap_{i=1}^r H_i$. Then $\text{cl}(t) = \cap_{i=1}^r \text{cl}(H_i)$.

Proof: Clearly $t \subseteq H_i$ implies $\text{cl}(t) \subseteq \text{cl}(H_i)$ for each i . Conversely, let $u \in t = \cap_i H_i$ (since t is nonempty), and take any $v \in \cap_i \text{cl}(H_i)$. Then we see that $(1 - \alpha)u + \alpha v \in H_i$ for each i and all $\alpha \in [0, 1]$. That is, $(1 - \alpha)u + \alpha v \in t$ for all $\alpha \in [0, 1]$. Taking $\alpha \rightarrow 1$, we see that v is a limit point of t . Thus, $\cap_i \text{cl}(H_i) \subseteq \text{cl}(t)$. $\square$

Corollary 5 *If $t = \cap_{i=1}^r H_i$, then the boundary of t is*

$$\partial t = \text{cl}(t) \setminus t = \bigcup_{\emptyset \neq I \subseteq \{1, \dots, r\}} \left[\left(\bigcap_{i \in I} \partial H_i \right) \cap \left(\bigcap_{i \notin I} H_i \right) \right].$$

Proof: Write $\text{cl}(t) = \cap_i \text{cl}(H_i) = \cap_i (\partial H_i \cup H_i)$ and distribute on the right side, then remove $t = \cap_i H_i$ from both sides. $\square$

In particular, $\partial t \subseteq \cup_i \partial H_i$.

Lemma 6 *Let u be a fixed utility function, and let $t, t' \in T$ be distinct, nonadjacent types. Then there exist finitely many hyperplanes, each one passing through u , whose union contains $\text{cl}(t) \cap \text{cl}(t')$.*

Proof: Since t, t' are open and disjoint, neither one can intersect the closure of the other, so $\text{cl}(t) \cap \text{cl}(t') \subseteq \partial t \cap \partial t'$. Now suppose $t = \cap_{i=1}^r H_i$ and $t' = \cap_{i'=1}^{r'} H_{i'}$. Applying Corollary 5 to both t and t' , and then distributing the intersection operator, we get

$$\partial t \cap \partial t' = \bigcup_{\substack{\emptyset \neq I \subseteq \{1, \dots, r\} \\ \emptyset \neq I' \subseteq \{1, \dots, r'\}}} B(I, I'),$$

where

$$B(I, I') = \left(\bigcap_{i \in I} \partial H_i \right) \cap \left(\bigcap_{i \notin I} H_i \right) \cap \left(\bigcap_{i' \in I'} \partial H_{i'} \right) \cap \left(\bigcap_{i' \notin I'} H_{i'} \right).$$

It therefore suffices to show that each set $B(I, I')$ is contained in a hyperplane that passes through u . We may assume $B(I, I')$ is nonempty, since otherwise we are done.

Now, $B(I, I')$ is a relatively open subset of $P(I, I') = (\cap_{i \in I} \partial H_i) \cap (\cap_{i' \in I'} \partial H_{i'})$. The set $P(I, I')$ is an affine set, that is, an intersection of hyperplanes. If $P(I, I')$ is itself a hyperplane, then since $B(I, I') \subseteq \text{cl}(t) \cap \text{cl}(t')$, it follows that t and t' are adjacent. This contradicts the hypothesis. Therefore $P(I, I')$ is an affine set of dimension at most $m - 2$,

and it is a standard fact that we can then find a hyperplane containing both $P(I, I')$ and u . $\square$

Proof of Proposition 3: Let $u \in \mathbb{R}^m$ be any given utility function. We know from Corollary 5 and Lemma 6 that there exists a finite collection K of hyperplanes such that

- for each $t \in T$, ∂t is contained in the union of the hyperplanes in K ;
- if t, t' are distinct, nonadjacent types, and $w \in \text{cl}(t) \cap \text{cl}(t')$, then there exists a hyperplane in K passing through u and w .

Now suppose $u \in t$, for some type t , and let t' be any other type. Suppose the mechanism f satisfies the local incentive constraints. We wish to show that $u \cdot (f(t) - f(t')) \geq 0$.

Let K be the collection of hyperplanes defined above. Because t' is an open set, we can choose $v \in t'$ not lying on any of the hyperplanes in K .

As in the proof of Proposition 1, define $u_\alpha = (1 - \alpha)u + \alpha v$ for $\alpha \in [0, 1]$, and let $[u, v]$ denote the line segment $\{u_\alpha \mid \alpha \in [0, 1]\}$. Each hyperplane in K can contain at most one point of $[u, v]$: if a hyperplane contained more than one point, it would contain the entire segment and in particular would contain $u_1 = v$, contradicting the choice of v . This implies, in particular, that the union of the hyperplanes of K can contain only finitely many points of $[u, v]$.

Now, each type is an open, convex subset of Euclidean space; therefore, its intersection with $[u, v]$ must be an open, convex subset of $[u, v]$. The only such subsets are those of the form $u_J = \{u_\alpha \mid \alpha \in J\}$, where J is an open subinterval of $[0, 1]$; that is, J is of the form (γ, δ) , $[0, \delta)$, or $(\gamma, 1]$. So let $t_0, t_1, \dots, t_k$ be all the types that intersect $[u, v]$, and write $t_i \cap [u, v] = u_{J_i}$, for $i = 0, 1, \dots, k$, where J_i is an open subinterval of $[0, 1]$. Because the types are disjoint, it follows that the J_i must be disjoint. Hence, there is an unambiguous way of arranging them in increasing order; relabeling if necessary, we may assume $J_0, J_1, \dots, J_k$ are in order. Moreover, $u \in t$ implies $0 \in J_0$ and $t_0 = t$; $v \in t'$ implies $1 \in J_k$ and $t_k = t'$.

To summarize: the types that intersect the segment $[u, v]$ are $t_0, t_1, \dots, t_k$, where

$$\begin{aligned} t_0 \cap [u, v] &= u_{[0, \delta_0]}, \\ t_1 \cap [u, v] &= u_{(\gamma_1, \delta_1)}, \\ &\vdots \\ t_{k-1} \cap [u, v] &= u_{(\gamma_{k-1}, \delta_{k-1})}, \\ t_k \cap [u, v] &= u_{(\gamma_k, 1]}; \end{aligned}$$

moreover, $\delta_0 \leq \gamma_1, \delta_1 \leq \gamma_2, \dots, \delta_{k-1} \leq \gamma_k$; and $t_0 = t, t_k = t'$.

Next we show that in fact $\delta_i = \gamma_{i+1}$ for each $i = 0, \dots, k-1$. Suppose instead that $\delta_i < \gamma_{i+1}$. Recall that, by assumption, the union of the closures of all types in T is convex. Hence, it contains the entire segment $[u, v]$. So for any α with $\delta_i < \alpha < \gamma_{i+1}$, then, u_α must lie in the closure of some type $\hat{t}$. But by definition of δ_i and γ_{i+1} , such a point u_α cannot belong to $\hat{t}$ proper, so it belongs to $\partial\hat{t}$. In particular, each such u_α belongs to one of the hyperplanes in K . But there are infinitely many such u_α , and we already showed that the union of the hyperplanes in K can contain only finitely many points of $[u, v]$. This is a contradiction.

We have now established that $\delta_i = \gamma_{i+1}$ for each $i = 0, \dots, k-1$. On the other hand, $u_{\delta_i} \in \text{cl}(t_i)$, and $u_{\gamma_{i+1}} \in \text{cl}(t_{i+1})$. So, $u_{\delta_i} \in \text{cl}(t_i) \cap \text{cl}(t_{i+1})$. If t_i, t_{i+1} are not adjacent types, then some hyperplane of K passes through u_{δ_i} and u . This again contradicts the fact that each hyperplane of K can intersect $[u, v]$ only once. So t_i, t_{i+1} are adjacent; hence we have $(t_i, t_{i+1}) \in S$, and likewise $(t_{i+1}, t_i) \in S$.

Now we finally can repeat the argument from Proposition 1. Pick arbitrary $\alpha_i \in J_i$ for each $i = 0, \dots, k$. Then, $u_{\alpha_i} \in t_i$. The local incentive constraints (t_i, t_{i+1}) and (t_{i+1}, t_i) ensure that

$$\begin{aligned} u_{\alpha_i} \cdot (f(t_i) - f(t_{i+1})) &\geq 0, \\ u_{\alpha_{i+1}} \cdot (f(t_{i+1}) - f(t_i)) &\geq 0. \end{aligned}$$

Multiplying these two inequalities by α_{i+1} and α_i , respectively, and adding gives

$$[\alpha_{i+1}u_{\alpha_i} - \alpha_iu_{\alpha_{i+1}}] \cdot (f(t_i) - f(t_{i+1})) \geq 0.$$

But $\alpha_{i+1}u_{\alpha_i} - \alpha_i u_{\alpha_{i+1}} = (\alpha_{i+1} - \alpha_i)u$. Dividing by the positive number $\alpha_{i+1} - \alpha_i$ then gives us

$$u \cdot (f(t_i) - f(t_{i+1})) \geq 0.$$

Summing over $i = 0, \dots, k-1$ and telescoping gives

$$u \cdot (f(t) - f(t')) = u \cdot (f(t_0) - f(t_k)) \geq 0,$$

which is what was to be proved. □

A particular case of Proposition 3 is, of course, the result that has already been mentioned a couple times, that on the full ordinal type space, the local incentive constraints are sufficient. (In this case, the union of the closures of all types is simply all of $\mathbb{R}^m$.) Another situation where Proposition 3 applies is when T consists of all ordinal types that respect a given partial ordering on X . For example, Bogomolnaia and Moulin [13] consider an allocation problem with real objects and a null object; all types have the same preference ordering on the real objects, but rank the null object differently relative to the real objects.

We have pointed out that Proposition 3 leads to an immediate strengthening of the Gibbard-Satterthwaite theorem by relaxing the requirement of (dominant-strategy) incentive-compatibility to local incentive-compatibility. It is natural to ask if the result can be made stronger still by using even fewer incentive constraints. In fact, there is no sufficient set of incentive constraints respecting symmetry among the outcomes that is smaller than the set of local incentive constraints.

To state this more clearly, suppose σ is a permutation of the set $\{1, 2, \dots, m\}$. For any type t , define a type t^σ as follows: if $x_1, \dots, x_m$ is the ordering of outcomes such that $t = \{u \mid u(x_1) > \dots > u(x_m)\}$, then $t^\sigma = \{u \mid u(x_{\sigma(1)}) > \dots > u(x_{\sigma(m)})\}$. Let Σ_{adj} be the set of *adjacent transpositions* — permutations of $\{1, 2, \dots, m\}$ that consist of switching i and $i+1$, for some i . Then, the set Σ_{adj} has the property that $\{(t, t^\sigma) \mid t \in T, \sigma \in \Sigma_{adj}\}$ is a sufficient set of incentive constraints (because these are precisely the local incentive constraints we have been using). In fact, Σ_{adj} is a *minimal* set of permutations with this sufficiency property. That is, on the full ordinal type space T , the set of incentive constraints $S' = \{(t, t^\sigma) \mid t \in T, \sigma \in \Sigma'\}$, where Σ' is any proper subset of Σ_{adj} , is not sufficient, not even for deterministic mechanisms. Indeed, suppose that, for some i , the transposition that switches i and $i+1$ is missing from Σ' . Then, the deterministic mechanism that assigns to every ordinal type its $(i+1)$ st favorite outcome satisfies the incentive constraints S' , but this mechanism is not fully incentive-compatible.

3.3 Single-peaked preferences

Our results so far have focused on essentially convex type spaces. One of the most important type spaces, however, is not convex: that of single-peaked preferences.

Fix an ordering $x_1, \dots, x_m$ of the outcomes in X . A strict preference ordering $\succ$ over X is *single-peaked* if there exists some k such that, whenever $i < j \leq k$ or $i > j \geq k$, we have $x_j \succ x_i$. For our purposes, an ordinal type is *single-peaked* if it represents a single-peaked ordering.

Single-peaked preferences have been popular in social choice ever since Black's [11] observation that, in voting situations, the median of the voters' favorite outcomes is a Condorcet winner: it does not lose to any other outcome in a pairwise majority vote. Moreover, the voting rule that chooses the median favorite outcome is dominant-strategy incentive-compatible. Single-peaked preferences are also important in economic applications because single-peakedness is the same as quasiconcavity of the utility function (aside from issues of indifference). Moulin [34] characterizes the deterministic voting systems that are dominant-strategy incentive-compatible and tops-only, both with and without an additional requirement of anonymity, under single-peaked preferences. (Moulin assumes the outcome space is the whole real line, but his proofs carry through almost unchanged for a finite outcome space.) Ehlers, Peters, and Storcken [21] extend this work to probabilistic mechanisms. Sprumont [45] and Barberà, Jackson, and Neme [6] study efficient, incentive-compatible rationing of a good to consumers with single-peaked preferences over quantities.

The space of single-peaked ordinal types does not meet the convexity condition of Proposition 3: if $m = 3$, then define two utility functions u, v by

$$\begin{aligned} u(x_1) &= 4, & u(x_2) &= 1, & u(x_3) &= 0, \\ v(x_1) &= 0, & v(x_2) &= 1, & v(x_3) &= 4. \end{aligned}$$

Then u and v represent single-peaked orderings, but $(u + v)/2$ does not.

However, we still have the result:

Proposition 7 *Let X be fixed, with an ordering $x_1, \dots, x_m$ of its elements. On the space of single-peaked ordinal types, the set of local incentive constraints is sufficient.*

The argument is a slight extension of that used for Proposition 3. Given a single-peaked utility function u and a type t' , we can find some $v \in t'$ such that the segment connecting u and v lies within the single-peaked domain, even though this may not be true for every $v \in t'$, or even almost every v . This is illustrated in Figure 2, which again

considers the case of three outcomes; single-peaked types are shown in gray. In the figure, to show that $u \cdot (f(t) - f(t')) \geq 0$, we cannot use the segment from u to v since it passes outside of the single-peaked domain, but the segment from u to v' does work.

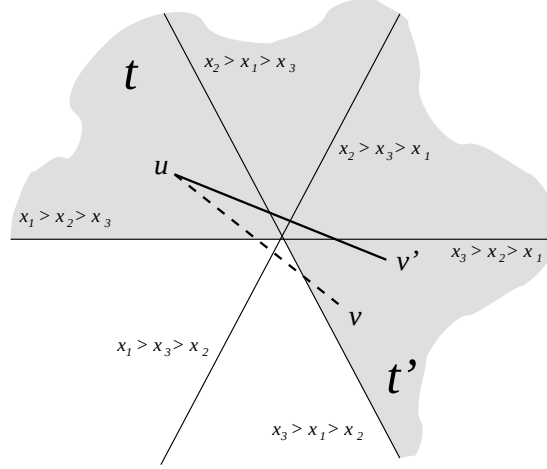


Figure 2: Argument for Proposition 7

Accordingly, suppose u is a given single-peaked utility function. Say that a utility function v is *good for* u if, for every $\alpha \in [0, 1]$, $(1 - \alpha)u + \alpha v$ lies in the closure of some single-peaked ordinal type. Note that as long as u is strict (i.e. gives different values to different outcomes), a sufficient condition for v to be good for u is that $(1 - \alpha)u + \alpha v$ be single-peaked whenever it is strict: the set $A_{ns} = \{\alpha \in [0, 1] \mid (1 - \alpha)u + \alpha v \text{ is not strict}\}$ is finite, so the closure of $[0, 1] \setminus A_{ns}$ is all of $[0, 1]$, and hence $(1 - \alpha)u + \alpha v$ is in the closure of some single-peaked ordinal type for each α .

Lemma 8 *Fix the single-peaked utility function u . For any ordinal type t' , there exists a nonempty open set contained in t' such that every v in the open set is good for u .*

Proof: We will first construct a v that is good for u . Let x_k be the outcome ranked highest by u , and let x_l be the outcome ranked highest by t' . If $l = k$, then any $v \in t'$ is good for u : since $u(x_i)$ and $v(x_i)$ are both increasing in i for $i \leq k$ and decreasing for $i \geq k$, the same is true of any weighted average $(1 - \alpha)u + \alpha v$, so each such weighted average is single-peaked (as long as it is strict).

Now suppose $l > k$ (the case $l < k$ is similar). So $u(x_i)$ is decreasing and $v(x_i)$ must be increasing in i for $k \leq i \leq l$. Choose $v(x_l)$ and $v(x_{l-1})$ arbitrarily, with $v(x_{l-1}) < v(x_l)$.

If $k < l - 1$ then successively choose $v(x_j)$ for $j = l - 2, l - 3, \dots, k$, such that

$$\frac{v(x_{j+2}) - v(x_{j+1})}{u(x_{j+1}) - u(x_{j+2})} < \frac{v(x_{j+1}) - v(x_j)}{u(x_j) - u(x_{j+1})}. \quad (1)$$

This can be done by choosing $v(x_j)$ low enough at each step. Finally, we can choose $v(x_j)$ for $j > l$ or $j < k$ so that v represents the ordering given by t' .

Now we will show that, for $\alpha \in [0, 1]$, $(1 - \alpha)u + \alpha v$ is single-peaked whenever it is strict. That is, we claim that $(1 - \alpha)u(x_i) + \alpha v(x_i)$ is increasing in i for i up to some peak p , and decreasing after that. Both $u(x_i)$ and $v(x_i)$ are increasing in i for $i \leq k$, and decreasing in i for $i \geq l$, so it suffices to focus on the range $k \leq i \leq l$. We will show that $(1 - \alpha)u(x_i) + \alpha v(x_i)$ cannot ever decrease and then increase within this range. That is, we will show that there is no $j \in \{k, \dots, l - 2\}$ such that

$$(1 - \alpha)u(x_j) + \alpha v(x_j) > (1 - \alpha)u(x_{j+1}) + \alpha v(x_{j+1}) \quad (2)$$

and

$$(1 - \alpha)u(x_{j+1}) + \alpha v(x_{j+1}) < (1 - \alpha)u(x_{j+2}) + \alpha v(x_{j+2}) \quad (3)$$

simultaneously hold; this will prove the claim.

Suppose (2) and (3) do both hold, for some j . (2) implies

$$\frac{1 - \alpha}{\alpha} > \frac{v(x_{j+1}) - v(x_j)}{u(x_j) - u(x_{j+1})}$$

while (3) implies

$$\frac{1 - \alpha}{\alpha} < \frac{v(x_{j+2}) - v(x_{j+1})}{u(x_{j+1}) - u(x_{j+2})}.$$

(Note we used the fact that $u(x_j) > u(x_{j+1}) > u(x_{j+2})$ to make sure the signs don't switch when we divide. We know $\alpha > 0$ since (3) is violated at $\alpha = 0$.) Combining these two inequalities gives a contradiction of (1), completing the proof of the claim.

At this point we have shown that any $v \in t'$ satisfying the inequalities (1) is good for u . Since these inequalities carve out a nonempty open subset of t' , the lemma is proven. $\square$

Proof of Proposition 7: Suppose that T is the space of single-peaked ordinal types and S the set of local incentive constraints. Let u be any utility function. As in the proof of Proposition 3, there exists a finite collection of hyperplanes K whose union contains the boundary of every type in T , and such that if t, t' are distinct types such that $(t, t') \notin S$,

and $w \in \text{cl}(t) \cap \text{cl}(t')$, then some hyperplane in K passes through u and w .

Now suppose $u \in t$ for some $t \in T$, and let $t' \in T$ be any other type. We wish to show that $u \cdot (f(t) - f(t')) \geq 0$.

By Lemma 8, there is a nonempty open set of types that are good for u , contained in t' . In particular, we can choose a $v \in t'$ that is good for u and does not lie on any of the hyperplanes in K . The fact that v is good for u means exactly that the segment $[u, v] = \{(1 - \alpha)u + \alpha v \mid \alpha \in [0, 1]\}$ is contained in the union of the closures of types in T .

From here the rest of the proof is a repeat of the argument used to prove Proposition 3; we again see that $u \cdot (f(t) - f(t')) \geq 0$. $\square$

It should be clear that this same proof applies also to the space of single-troughed ordinal types. (A preference ordering on X is *single-troughed*, or *single-dipped*, if it is the reverse of a single-peaked ordering: for some k , $i < j \leq k$ or $i > j \geq k$ imply $x_i \succ x_j$ [30].)

In addition, single-peaked preferences on a linear ordering of X can be generalized to single-peaked preferences on a tree [19, 18]. Suppose we are given an (undirected) tree whose vertices are the outcomes in X ; then a strict preference ordering $\succ$ on X is single-peaked if there exists some x such that, for all distinct y, z , if y lies on the path from x to z in the tree, then $y \succ z$. Classical single-peaked preferences are just the special case where the tree is a path. It is straightforward to check that Lemma 8, and so also Proposition 7, can be generalized to the space of single-peaked ordinal types on any given tree. (One can also define single-troughed preferences on a tree in the obvious way, and the results hold there as well.)

3.4 Single-crossing preferences

Besides single-peaked preferences, another economically important class of ordinal type spaces is given by single-crossing preferences. These are defined as follows: Fix an ordering $x_1, \dots, x_m$ of the elements of X . A sequence $\succ_1, \dots, \succ_p$ of distinct strict preference orderings is a *single-crossing preference domain* if the following holds: whenever $i < j$ and $x_j \succ_k x_i$ for some k , we also have $x_j \succ_l x_i$ for all $l > k$. Single-crossing ordinal preferences arise in some economic models, such as the redistributive taxation models of Roberts [36] and Meltzer and Richard [33] (see [41] for references to other applications). Just as with single-peaked preferences, preferences coming from any single-crossing domain satisfy a median voter property — the favorite outcome of the voter with the median preference is a Condorcet winner, and the voting scheme that chooses this outcome is dominant-strategy

incentive-compatible. More generally, Saporiti [41] characterizes the voting schemes that are anonymous, unanimous, and dominant-strategy incentive-compatible on a maximal single-crossing preference domain.

For any strict preference ordering $\succ$ on X , let $V(\succ) = \{(i, j) \mid i < j, x_j \succ x_i\}$. By definition, a sequence of preference orderings $\succ_1, \dots, \succ_p$ is a single-crossing preference domain if and only if $V(\succ_1) \subseteq V(\succ_2) \subseteq \dots \subseteq V(\succ_p)$. In fact, these inclusions must all be strict, since any ordering $\succ$ can be uniquely reconstructed from $V(\succ)$. Therefore we have $|V(\succ_1)| < \dots < |V(\succ_p)|$. Call the domain a *successive single-crossing preference domain* if $|V(\succ_{k+1})| = |V(\succ_k)| + 1$ for each $k = 1, \dots, p-1$.

Notice that if we are given a single-crossing preference domain that is not successive, we can add more preference orderings to make it successive. For suppose that $|V(\succ_{k+1})| - |V(\succ_k)| > 1$ for some k . There must be some two alternatives x_i, x_j that are ranked consecutively by $\succ_k$, with $x_i \succ_k x_j$ but $x_j \succ_{k+1} x_i$. (Otherwise, $\succ_{k+1}$ would agree with $\succ_k$ for every pair of outcomes ranked consecutively by $\succ_k$, implying $\succ_{k+1} = \succ_k$.) Therefore $i < j$, $(i, j) \notin V(\succ_k)$ and $(i, j) \in V(\succ_{k+1})$. By switching the positions of x_i and x_j in $\succ_k$, we get a new ordering $\succ'$ with $V(\succ') = V(\succ_k) \cup \{(i, j)\}$, and hence $V(\succ_k) \subset V(\succ') \subset V(\succ_{k+1})$. This means that $\succ_1, \dots, \succ_k, \succ', \succ_{k+1}, \dots, \succ_p$ is again a single-crossing preference domain. Repeatedly filling in gaps in the sequence of values $|V(\succ_k)|$ in this way, we eventually obtain a successive single-crossing preference domain. In particular, any maximal single-crossing preference domain (the domains studied in [41]) must be successive.

Call an ordinal type space $T = \{t_1, \dots, t_p\}$ a *successive single-crossing ordinal type space* if each t_k is the ordinal type corresponding to $\succ_k$, for some successive single-crossing preference domain $\succ_1, \dots, \succ_p$. In this case, $\succ_k$ and $\succ_{k+1}$ differ by a switch of two consecutive outcomes, for each k , while $\succ_k$ and $\succ_l$ do not for $|k - l| > 1$; hence the local incentive constraints are precisely those of the form (t_k, t_{k+1}) or (t_{k+1}, t_k) . We shall show that on such a type space, the local incentive constraints are sufficient. This result may be surprising, since these incentive constraints are especially parsimonious — each type is explicitly prevented from imitating just two other types.

Proposition 9 *On any successive single-crossing ordinal type space, the local incentive constraints are sufficient.*

The strategy of proof is a little different from that used for the previous propositions. Instead of connecting any given utility function u to any other type t' by a single line segment and breaking it into short pieces, we find a sequence of parallel line segments,

each connecting two consecutive types t_k, t_{k+1} , but such that each segment need not begin where the previous one ended.

Proof: Suppose the mechanism f satisfies the local incentive constraints. Fix any two types $t_l, t_{l'}$, and let $u \in t_l$. We wish to show that $u \cdot (f(t_l) - f(t_{l'})) \geq 0$. We will show this for $l' > l$; the proof for $l' < l$ is similar.

In fact it suffices to show that

$$u \cdot (f(t_k) - f(t_{k+1})) \geq 0 \quad \text{for } k \geq l, \quad (4)$$

since then we can sum up (4) for $k = l, l+1, \dots, l'-1$ to obtain $u \cdot (f(t_l) - f(t_{l'})) \geq 0$.

So fix $k \geq l$, and also define $M = \max_x u(x) - \min_x u(x)$. Write $V(\succ_{k+1}) \setminus V(\succ_k) = \{(i, j)\}$ by successiveness; then $i < j$, and $\succ_k$ ranks x_i just above x_j . Because $u \in t_l$ with $l \leq k$, single-crossing implies that $u(x_i) > u(x_j)$ also. Let v be any utility function representing $\succ_k$ such that $v(x_i) - v(x_j) < u(x_i) - u(x_j)$, and $|v(x) - v(y)| > M$ for all distinct outcomes $x, y \in X$ other than x_i and x_j . Because $\succ_k$ ranks x_i and x_j consecutively, we can do this. Then, $v - u$ is consistent with $\succ_{k+1}$. Check: we have $v(x_i) - u(x_i) < v(x_j) - u(x_j)$, so that $v - u$ ranks x_i below x_j . And for any other pair x, y , we have $|v(x) - v(y)| > |u(x) - u(y)|$, so $[v(x) - u(x)] - [v(y) - u(y)]$ has the same sign as $v(x) - v(y)$. That is, $v - u$ ranks all pairs of outcomes in the same order as v does, except $\{x_i, x_j\}$. Since $V(\succ_{k+1}) = V(\succ_k) \cup \{(i, j)\}$, this means that $v - u$ represents $\succ_{k+1}$.

So, $v \in t_k$ and $v - u \in t_{k+1}$. The local incentive constraints then give

$$\begin{aligned} v \cdot (f(t_k) - f(t_{k+1})) &\geq 0, \\ [v - u] \cdot (f(t_{k+1}) - f(t_k)) &\geq 0. \end{aligned}$$

Adding these two gives exactly (4), and this completes the proof. $\square$

The hypothesis of successiveness in Proposition 9 cannot be dropped, even if the set of local incentive constraints is modified in the natural way. That is, it is not the case that, for any single-crossing ordinal type space $\{t_1, \dots, t_p\}$ (corresponding to a single-crossing preference domain $\succ_1, \dots, \succ_p$), the set consisting of the incentive constraints (t_k, t_{k+1}) and (t_{k+1}, t_k) , for $1 \leq k < p$, is sufficient. For a counterexample, consider the following

three orderings:

$$\begin{aligned}\succ_1: & \quad x_1 \succ_1 x_2 \succ_1 x_3 \succ_1 x_4; \\ \succ_2: & \quad x_2 \succ_2 x_1 \succ_2 x_3 \succ_2 x_4; \\ \succ_3: & \quad x_4 \succ_3 x_2 \succ_3 x_1 \succ_3 x_3,\end{aligned}$$

and the corresponding ordinal types t_1, t_2, t_3 . Let f be the mechanism mapping the types to lotteries over (x_1, x_2, x_3, x_4) as follows:

$$f(t_1) = (1/4, 1/4, 1/2, 0); \quad f(t_2) = (0, 1/2, 1/2, 0); \quad f(t_3) = (1/2, 0, 0, 1/2).$$

Then f satisfies the incentive constraints $(t_1, t_2), (t_2, t_1), (t_2, t_3), (t_3, t_2)$, but not (t_1, t_3) , so it is not fully incentive-compatible.

3.5 Mechanisms with monetary transfers

So far we have focused on mechanisms whose result is a lottery over a finite set of outcomes. However, much of the field of mechanism design concerns settings where monetary transfers are available: agents have quasilinear utility in outcomes and money, and a mechanism specifies both an outcome (or lottery over outcomes) and a monetary transfer. This is the case, for example, in much of the literature on auctions and public projects, as well as the screening literature in contract theory. This field is too large to survey here (see [31, ch. 23] for a general introduction), but it seems worthwhile to mention how our ideas apply in this context. In fact, the approach of Subsection 3.1 carries over easily to settings with transfers and quasilinear utility. We will focus on cardinal type spaces, as is standard in this literature.

The definitions of an outcome space and a cardinal type space are as before, but for purposes of this subsection, a *mechanism* on a type space T now consists of an ordered pair (f, p) , where $f : T \rightarrow \Delta(X)$ as before, and $p : T \rightarrow \mathbb{R}$ is a *payment function*. The interpretation is that when the agent announces type u , he receives the lottery over outcomes $f(u)$ and also receives the net monetary transfer $p(u)$. (The monetary transfer may be uncertain, in which case $p(u)$ is its expected value.) If his true utility function is u and he announces v , then, the payoff he receives is $u \cdot f(v) + p(v)$. We therefore say that the mechanism *satisfies* the incentive constraint (u, v) if

$$u \cdot f(u) + p(u) \geq u \cdot f(v) + p(v),$$

or equivalently

$$u \cdot (f(u) - f(v)) + (p(u) - p(v)) \geq 0.$$

As before, a mechanism satisfies a set S of incentive constraints if it satisfies every constraint in the set. S is *sufficient* if every mechanism that satisfies S is fully incentive-compatible, that is, satisfies every possible constraint $(u, v) \in T \times T$.

In this setting, sets of local incentive constraints are defined as before. We then have:

Proposition 10 *In the setting of mechanisms with monetary transfers, on any convex cardinal type space T , any set of local incentive constraints is sufficient.*

Proof: Let S be a set of local incentive constraints, and (f, p) a mechanism satisfying S . We wish to show that $u \cdot (f(u) - f(v)) + (p(u) - p(v)) \geq 0$ for every $u, v \in T$.

So consider any such u, v . Write $u_\alpha = (1 - \alpha)u + \alpha v$, for $\alpha \in [0, 1]$. As in the proof of Proposition 1, we are assured the existence of a finite sequence $0 = \alpha_0 < \alpha_1 < \dots < \alpha_k = 1$ with $(u_{\alpha_i}, u_{\alpha_{i+1}}) \in S$ and $(u_{\alpha_{i+1}}, u_{\alpha_i}) \in S$ for each $i = 0, 1, \dots, k-1$. These local incentive constraints then give us

$$\begin{aligned} u_{\alpha_i} \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) + (p(u_{\alpha_i}) - p(u_{\alpha_{i+1}})) &\geq 0, \\ u_{\alpha_{i+1}} \cdot (f(u_{\alpha_{i+1}}) - f(u_{\alpha_i})) + (p(u_{\alpha_{i+1}}) - p(u_{\alpha_i})) &\geq 0. \end{aligned}$$

Multiplying these two constraints by α_{i+1} and α_i , respectively, and adding gives

$$[\alpha_{i+1}u_{\alpha_i} - \alpha_iu_{\alpha_{i+1}}] \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) + (\alpha_{i+1} - \alpha_i)(p(u_{\alpha_i}) - p(u_{\alpha_{i+1}})) \geq 0.$$

But again, $\alpha_{i+1}u_{\alpha_i} - \alpha_iu_{\alpha_{i+1}} = (\alpha_{i+1} - \alpha_i)u$, so we have

$$(\alpha_{i+1} - \alpha_i)u \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) + (\alpha_{i+1} - \alpha_i)(p(u_{\alpha_i}) - p(u_{\alpha_{i+1}})) \geq 0.$$

Dividing by the positive number $\alpha_{i+1} - \alpha_i$ now gives

$$u \cdot (f(u_{\alpha_i}) - f(u_{\alpha_{i+1}})) + (p(u_{\alpha_i}) - p(u_{\alpha_{i+1}})) \geq 0.$$

Finally, summing over $i = 0, 1, \dots, k-1$ and telescoping gives

$$u \cdot (f(u_{\alpha_0}) - f(u_{\alpha_k})) + (p(u_{\alpha_0}) - p(u_{\alpha_k})) \geq 0,$$

or more simply put,

$$u \cdot (f(u) - f(v)) + (p(u) - p(v)) \geq 0.$$

□

3.6 On methods of proof

So far, all of our proofs of sufficiency results have followed the same general method: showing that the linear inequality corresponding to any desired incentive constraint can be obtained by “adding up” the inequalities corresponding to local incentive constraints. It is natural to ask whether or not every sufficiency result can always be proven in this manner. In this subsection, we investigate that question and show that the answer is negative: we give a type space and a set of local incentive constraints that are sufficient, but whose sufficiency cannot be proven by adding up. Along the way, we develop tools to study when an incentive constraint can be proven by adding up. In particular, we show (Proposition 13) that any such proof essentially must be of the form used to prove Proposition 1, by breaking up a single line segment into pieces. This result is of some independent interest, and it will be useful again in Section 4 when we discuss examples of insufficiency.

We must first be precise about what it means for an incentive constraint to be provable by adding up other constraints. We will again focus on cardinal type spaces for the rest of this subsection. Let T be a cardinal type space, and let S be a set of incentive constraints. Let $\mathbf{1} \in \mathbb{R}^m$ denote the vector all of whose components are 1, and let e_i denote the i th unit vector for $i = 1, \dots, m$. For any mechanism f , we have

$$\mathbf{1} \cdot f(u) = 1 \tag{5}$$

for all $u \in T$, and

$$e_i \cdot f(u) \geq 0 \tag{6}$$

for $i = 1, \dots, m$ and all $u \in T$. If f satisfies S , then we also have

$$u \cdot (f(u) - f(v)) \geq 0 \tag{7}$$

for each $(u, v) \in S$.

We say that an incentive constraint $(u^*, v^*) \in T \times T$ is *provable from S by adding up*

if the inequality

$$u^* \cdot (f(u^*) - f(v^*)) \geq 0 \quad (8)$$

can be obtained as a finite linear combination of the equations (5) and inequalities (6), (7), with nonnegative coefficients on the inequalities. That is, (u^*, v^*) is provable from S by adding up if there exist real numbers

- a_u for $u \in T$,
- b_{iu} for $i = 1, \dots, m$, $u \in T$, and
- c_{uv} for $(u, v) \in S$,

such that all but finitely many of these numbers are zero, all the b_{iu} and c_{uv} are nonnegative, and such that adding up a_u times (5), b_{iu} times (6), and c_{uv} times (7) gives (8). (For notational convenience, we will assume c_{uv} to be defined for all $u, v \in T$, with $c_{uv} = 0$ whenever $(u, v) \notin S$.)

We can write out the adding-up conditions explicitly, by comparing coefficients of $f(u)$, for each $u \in T$. Assume $u^* \neq v^*$ (otherwise (8) just reads $0 = 0$ which is trivially provable by adding up). Then the adding-up condition says that for each u , we have

$$a_u \mathbf{1} + \sum_{i=1}^m b_{iu} e_i + \sum_{v \in T} c_{uv} u - \sum_{v \in T} c_{vu} v = \begin{cases} u^* & \text{if } u = u^* \\ -u^* & \text{if } u = v^* \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Also, for the constant terms, the adding-up condition is simply

$$\sum_{u \in T} a_u = 0. \quad (10)$$

We say that the set S of incentive constraints *implies* the incentive constraint $(u^*, v^*) \in T \times T$ if every mechanism that satisfies S also satisfies (u^*, v^*) .

The present question is: If S implies (u^*, v^*) , must the constraint (u^*, v^*) necessarily be provable from S by adding up? When S is *finite*, the answer is affirmative:

Lemma 11 *If T is a cardinal type space and S a finite set of incentive constraints that implies the incentive constraint (u^*, v^*) , then (u^*, v^*) is provable from S by adding up.*

Proof: We may as well assume that T consists only of the types that appear in constraints of S and u^*, v^* , since any other types that may exist are irrelevant. Thus,

T is finite. A mechanism f satisfying S then consists simply of a choice of $m \cdot |T|$ real numbers — the components of the $|T|$ vectors $f(u)$ for $u \in T$ — satisfying (5), (6), and also (7) for $(u, v) \in S$. The hypothesis is that any such numbers must also satisfy (8).

This can be recast as a linear programming statement: for any choice of $m \cdot |T|$ real numbers satisfying the nonnegativity constraints (6) and the linear equations (5) and inequalities (7), the minimum value of the linear function $u^* \cdot (f(u^*) - f(v^*))$ is 0. (Note that this minimum is attained, for example, by any mechanism such that $f(u)$ is constant across all u .) The duality theorem of linear programming then tells us that (8) is expressible as a linear combination of (5), (6), (7), with nonnegative coefficients on the inequalities. That is, (u^*, v^*) is provable from S by adding up. $\square$

In particular, this result implies that on any finite cardinal type space, if some set of incentive constraints is sufficient, then sufficiency can be proven by adding up.

(With a little more work, we could prove an analogue of Lemma 11 for polyhedral type spaces. However, we will not bother to present it here, since our interest in provability by adding up is merely instrumental — no exhaustive investigation is intended.)

To proceed further, it will be helpful to have an alternative, cleaner definition of provability by adding up. Let $\Pi \subseteq \mathbb{R}^m$ be the hyperplane orthogonal to $\mathbf{1}$, and for any $u \in \mathbb{R}^m$, let $\bar{u}$ denote its orthogonal projection onto Π .

Lemma 12 *Assume $u^* \neq v^*$. Then (u^*, v^*) is provable from S by adding up if and only if there exist numbers $c_{uv} \geq 0$, finitely many of which are nonzero, such that the equation*

$$\sum_{v \in T} c_{uv} \bar{u} - \sum_{v \in T} c_{vu} \bar{v} = \begin{cases} \bar{u}^* & \text{if } u = u^* \\ -\bar{u}^* & \text{if } u = v^* \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

holds for each $u \in T$, and $c_{uv} = 0$ unless $(u, v) \in S$.

Proof: First suppose that (u^*, v^*) is provable from S by adding up under the original definition; let a_u, b_{iu}, c_{uv} be the coefficients satisfying (9). By summing (9) over all choices of u we get $\sum_u a_u \mathbf{1} + \sum_i \sum_u b_{iu} e_i = 0$. (On the left side, each c_{uv} occurs once multiplied by u and once multiplied by $-u$. On the right side, we get one u^* , one $-u^*$, and all zeroes otherwise.) From (10), this reduces to $\sum_i \sum_u b_{iu} e_i = 0$. Since the b_{iu} are nonnegative, they must all be zero. Once we know this, then, taking (9) and projecting orthogonally onto Π gives (11).

Conversely, suppose there are coefficients c_{uv} satisfying (11). Put $b_{iu} = 0$ for all i and

all u . Note that (11) implies that for each u , the expression

$$\begin{cases} \sum_v c_{uv}u - \sum_v c_{vu}v - u^* & (u = u^*) \\ \sum_v c_{uv}u - \sum_v c_{vu}v + u^* & (u = v^*) \\ \sum_v c_{uv}u - \sum_v c_{vu}v & (\text{otherwise}) \end{cases}$$

must be some multiple of $\mathbf{1}$. Choose a_u so that this expression is equal to $-a_u\mathbf{1}$. Then it is immediate that (9) is satisfied for each u . Moreover, summing (9) across all $u \in T$, the c_{uv} terms cancel as in the previous paragraph, and we are simply left with $\sum_u a_u\mathbf{1} = 0$; hence, with this choice of a_u , (10) is satisfied as well. Finally, $a_u \neq 0$ only when $u = u^*, v^*$ or when c_{uv} or c_{vu} is nonzero for some v ; thus, only finitely many of the a_u are nonzero. Thus, the original definition of provability by adding up is satisfied. $\square$

Now we arrive at a result that is crucial in understanding the limits of provability by adding up. For $u^*, v^* \in T$, let $T_{[u^*, v^*]}$ be the set of all types in T that are equivalent to $(1 - \alpha)u^* + \alpha v^*$ for some $\alpha \in [0, 1]$, and let

$$S_{[u^*, v^*]} = \{(u, v) \in S \mid u, v \in T_{[u^*, v^*]}\}.$$

Proposition 13 *Let T be a cardinal type space and S a set of incentive constraints such that (u^*, v^*) is provable from S by adding up. Assume that v^* is not equivalent to $-u^*$. Then (u^*, v^*) is provable from $S_{[u^*, v^*]}$ by adding up.*

This result says that if an incentive constraint (u^*, v^*) can be proved by adding up constraints in S , then it can be proved by adding up in a way that only uses types equivalent to convex combinations of u^* and v^* — the kind of argument used to prove Proposition 1. The proof of Proposition 13 proceeds by taking the coefficients c_{uv} satisfying (11) and successively replacing them by zero, checking that (11) still holds at each step, until only constraints in $S_{[u^*, v^*]}$ have nonzero coefficients.

Proof: We may assume that u^* is not indifferent, since otherwise the conclusion is immediate: (11) holds with all c_{uv} equal to 0. We also assume $u^* \neq v^*$; otherwise the conclusion is again trivial.

Let c_{uv} be the coefficients satisfying (11), with $c_{uv} > 0$ only if $(u, v) \in S$. We may as well assume that S consists only of the (finitely many) incentive constraints (u, v) for which $c_{uv} > 0$, and T consists only of the types appearing in these constraints — all other types can be ignored.

Now consider any fixed vector $w \in \Pi$ with the following properties:

- (i) $w \cdot \bar{u}^* > 0$;
- (ii) $w \cdot \bar{v}^* \geq 0$;
- (iii) if $u \in T$ and $w \cdot \bar{u} = 0$, then $\bar{u} = 0$.

We claim that if $(u, v) \in S$ such that either

- (a) $w \cdot \bar{u} > 0$ and $w \cdot \bar{v} < 0$, or
- (b) $w \cdot \bar{u} > 0$ and $w \cdot \bar{v} = 0$ and $v \neq v^*$, or
- (c) $w \cdot \bar{u} < 0$ and $w \cdot \bar{v} \geq 0$,

then $c_{uv} = 0$.

Proof: Consider any $u \in T$ such that $w \cdot \bar{u} < 0$. Take the dot product of w with (11). We get

$$\sum_{v \in T} c_{uv}(w \cdot \bar{u}) - \sum_{v \in T} c_{vu}(w \cdot \bar{v}) = 0$$

(note that $u \neq u^*, v^*$). Now sum over all u such that $w \cdot \bar{u} < 0$. For each incentive constraint $(u, v) \in S$ such that $w \cdot \bar{u} < 0$ and $w \cdot \bar{v} < 0$, the term $c_{uv}(w \cdot \bar{u})$ appears once with a + sign and once with a - sign, so these cancel out. The remaining terms give us

$$\sum_{w \cdot \bar{u} < 0; w \cdot \bar{v} \geq 0} c_{uv}(w \cdot \bar{u}) - \sum_{w \cdot \bar{u} < 0; w \cdot \bar{v} \geq 0} c_{vu}(w \cdot \bar{v}) = 0.$$

Since each c_{uv} is nonnegative, every term in the first sum is ≤ 0 and every term in the second sum is ≥ 0 . Hence, every term must be equal to zero. This implies that whenever $w \cdot \bar{u} < 0$ and $w \cdot \bar{v} \geq 0$, $c_{uv} = 0$, and when moreover $w \cdot \bar{v} > 0$, we also have $c_{vu} = 0$.

This covers (a) and (c). For (b), when $w \cdot \bar{v} = 0$ and $v \neq v^*$, (11) for v gives $\sum_u c_{vu}\bar{v} - \sum_u c_{uv}\bar{u} = 0$. Dotting with w gives $\sum_u c_{uv}(w \cdot \bar{u}) = 0$ (after canceling). We have already established that $c_{uv} = 0$ if $w \cdot \bar{u} < 0$, so all the terms on the left are nonnegative, and hence they must all be zero. So $c_{uv} = 0$ whenever $w \cdot \bar{u} > 0$.

This proves the claim.

Next, for each $u, v \in T$, define $c'_{uv} = c_{uv}$ if $w \cdot \bar{u} \geq 0$ and $w \cdot \bar{v} \geq 0$; and $c'_{uv} = 0$ otherwise. Then we again have, for each u ,

$$\sum_v c'_{uv}\bar{u} - \sum_v c'_{vu}\bar{v} = \begin{cases} \bar{u}^* & \text{if } u = u^* \\ -\bar{u}^* & \text{if } u = v^* \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

Proof: If u is such that $w \cdot \bar{u} < 0$ then (12) is trivial since both sides are zero. If $w \cdot \bar{u} > 0$, then the left side of (12) differs from the left side of (11) by the terms $c_{uv}\bar{u}$ and $-c_{vu}\bar{v}$ for $w \cdot \bar{v} < 0$. These are all zero, by cases (a) and (c) of the claim, respectively; thus (12) follows from (11). If $w \cdot \bar{u} = 0$ and $u \neq v^*$, then again all the left-hand-side terms of (12) are zero:

- all the $c'_{uv}\bar{u}$ are zero because $\bar{u} = 0$, by condition (iii) on w ;
- $c'_{vu}\bar{v} = 0$ for $w \cdot \bar{v} > 0$ by (b) of the claim;
- $c'_{vu}\bar{v} = 0$ for $w \cdot \bar{v} = 0$ again by (iii) on w ;
- $c'_{vu}\bar{v} = 0$ for $w \cdot \bar{v} < 0$ by definition of c'_{vu} .

So both sides of (12) are zero, and it again holds.

Thus, (12) is verified for all u except possibly for $u = v^*$. But summing (12) over all $u \in T$ gives the identity $0 = 0$; so if it holds for all u except $u = v^*$, it must hold for $u = v^*$ as well.

At this point, we have shown the following: If we start with coefficients c_{uv} for which (11) holds, pick any $w \in \Pi$ satisfying (i)-(iii), and replace c_{uv} by 0 whenever $w \cdot \bar{u} < 0$ or $w \cdot \bar{v} < 0$, then (11) still holds.

If we find any finite set of vectors $w_1, \dots, w_q \in \Pi$, each satisfying conditions (i)-(iii), and for each w_i we successively replace c_{uv} by 0 whenever $w_i \cdot \bar{u} < 0$ or $w_i \cdot \bar{v} < 0$, then the resulting coefficients will still satisfy (11).

Now let $T_{[u^*, v^*]+}$ consist of the types in $T_{[u^*, v^*]}$ together with all indifferent types (equivalently, all types that are equivalent to $\alpha u^* + \beta v^*$ for some $\alpha, \beta \geq 0$); and let $S_{[u^*, v^*]+} = \{(u, v) \in S \mid u, v \in T_{[u^*, v^*]+}\}$. We will show that, for any $u \in T$ that is not in $T_{[u^*, v^*]+}$, there is some $w \in \Pi$ satisfying (i)-(iii) with $w \cdot \bar{u} < 0$. If we consider each such w in turn, and successively replace c_{uv} 's by 0 as in the previous paragraph, we will be left with coefficients $c_{uv} \geq 0$ that still satisfy (11), and such that $c_{uv} = 0$ unless $u, v \in T_{[u^*, v^*]+}$. Therefore, we will have shown that (u, v) is provable from $S_{[u^*, v^*]+}$ by adding up.

Thus, consider any $u \in T \setminus T_{[u^*, v^*]+}$. We wish to show that there exists $w \in \Pi$ satisfying (i)-(iii) with $w \cdot \bar{u} < 0$. The assumptions that v^* is not equivalent to $-u^*$ and u^* is not indifferent imply that there exists $w' \in \Pi$ with

$$w' \cdot \bar{u}^* > 0, \quad w' \cdot \bar{v}^* \geq 0$$

and the latter inequality holding strictly unless $\bar{v}^* = 0$. The assumption $u \notin T_{[u^*, v^*]+}$ implies that $\bar{u}$ is not a nonnegative combination of $\bar{u}^*$ and $\bar{v}^*$; hence there is some $w'' \in \Pi$ such that

$$w'' \cdot \bar{u}^* \geq 0, \quad w'' \cdot \bar{v}^* \geq 0, \quad w'' \cdot \bar{u} < 0.$$

Taking $w = w' + \kappa w''$ for large κ will give (i), (ii), and $w \cdot \bar{u} < 0$. Finally, by perturbing w slightly, we can ensure $w \cdot \bar{v} \neq 0$ for all $v \in T$, $\bar{v} \neq 0$, without breaking any of the strict inequalities; thus we get (iii) as well.

At this point we have finished showing that (u^*, v^*) is provable from $S_{[u^*, v^*]+}$ by adding up.

If v^* is indifferent, then $S_{[u^*, v^*]+} = S_{[u^*, v^*]}$ and so we are done. Otherwise, we have to do just a little more work.

Let c_{uv} now be the coefficients used to prove (u^*, v^*) from $S_{[u^*, v^*]+}$ by adding up (i.e. the coefficients satisfying (11)). Whenever $\bar{u} = 0$, we can replace c_{uv} by 0 without affecting the validity of (11) (since c_{uv} only ever appears as part of the product $c_{uv}\bar{u}$). So we may assume $c_{uv} = 0$ whenever u is indifferent.

Since u^*, v^* are both non-indifferent and v^* is not equivalent to $-u^*$, we can find $w \in \Pi$ such that $w \cdot \bar{u}^* > 0$ and $w \cdot \bar{v}^* > 0$. Thus, for any element of $T_{[u^*, v^*]+}$ that is not indifferent, its projection has positive dot product with w .

Now for any indifferent u , taking (11) and dotting with w gives $-\sum_v c_{vu}(w \cdot \bar{v}) = 0$. Each term in the sum is nonnegative, so they must all be zero. Hence $c_{vu} = 0$ whenever $\bar{v}$ has positive dot product with w ; and the remaining $v \in T_{[u^*, v^*]+}$ are indifferent, so $c_{vu} = 0$ for them too by assumption. Thus, if u is indifferent then $c_{uv}, c_{vu} = 0$ for all v .

But this means that (11) holds with c_{uv} zero unless $u, v \in S_{[u^*, v^*]}$, so in fact (u^*, v^*) is provable from $S_{[u^*, v^*]}$ by adding up.

□

Proposition 13 will be useful because it provides a tool to show that a constraint is *not* provable by adding up: (u^*, v^*) cannot be provable by adding up unless it is provable in a very specific such way.

With this tool, we can carry out the promise made at the beginning of this subsection — to give an example of a type space and a set of local incentive constraints that are sufficient, but whose sufficiency cannot be proven by adding up. In fact, we will give a type space such that *any* set of local incentive constraints is sufficient, yet there exist fairly large such sets whose sufficiency cannot be proven by adding up.

Let X have four elements, and let w be some utility function on X that is not indif-

ferent. Let T_w be the set of all cardinal types that are either indifferent or equivalent to w , and let $T = \mathbb{R}^4 \setminus T_w$ be the set of cardinal types not in T_w . Say that two types $u, v \in T$ are T_w -opposed if there exists $\alpha \in (0, 1)$ such that $(1 - \alpha)u + \alpha v \in T_w$. Let S be any set of local incentive constraints such that if u and v are T_w -opposed, then $(u, v) \notin S$.

This requirement on S can be easily satisfied. Indeed, for each $u \in T$, start with any neighborhood N_u , and let $d(u, T_w) > 0$ be the Euclidean distance from u to T_w . Then the set $N'_u = \{v \in N_u \mid d(u, v) < d(u, T_w)\}$ is again an open neighborhood of u , not containing any types T_w -opposed to u . So $S = \{(u, v) \mid u \in N'_v \text{ or } v \in N'_u\}$ is a set of local incentive constraints as above.

Proposition 14 *With T, S as above, S is sufficient. However, if we choose $u^*, v^* \in T$ that are T_w -opposed, with u^* not equivalent to $-v^*$, then (u^*, v^*) is not provable from S by adding up.*

Proof: First we show that S is sufficient. Let f be any mechanism that satisfies S . For any possible incentive constraint (u, v) , if u and v are not T_w -opposed, then the entire line segment from u to v is contained in T . Therefore, exactly the same argument as in the proof of Proposition 1 shows that f satisfies (u, v) .

So we need only deal with the case where u, v are T_w -opposed. In this case, notice that we can choose $u_i \in T$ arbitrarily close to $(u + v)/2$ such that u_i is not T_w -opposed to either u or v . (Any type T_w -opposed to u must lie on the hyperplane Π_{uw} generated by u, w , and $\mathbf{1}$. Similarly, any type T_w -opposed to v must lie on the hyperplane generated by $v, w, \mathbf{1}$, which is again Π_{uw} . There are types in T arbitrarily close to $(u + v)/2$ not lying on this hyperplane.) For any such u_i , then, we have already shown that f satisfies the constraints $(u, u_i), (v, u_i), (u_i, v)$; that is:

$$u \cdot (f(u) - f(u_i)) \geq 0, \tag{13}$$

$$v \cdot (f(v) - f(u_i)) \geq 0, \tag{14}$$

$$u_i \cdot (f(u_i) - f(v)) \geq 0. \tag{15}$$

So we can choose a sequence of types $u_1, u_2, \dots$ in T with $u_i \rightarrow (u + v)/2$, such that (13)-(15) are satisfied for each u_i . Moreover, because the image of f is contained in the compact set $\Delta(X)$, we may assume by passing to a subsequence that $f(u_i)$ converges to

some limit f^* . Then, taking limits, we get

$$u \cdot (f(u) - f^*) \geq 0, \quad (16)$$

$$v \cdot (f(v) - f^*) \geq 0, \quad (17)$$

$$\frac{u+v}{2} \cdot (f^* - f(v)) \geq 0. \quad (18)$$

Adding (16), (17), and twice (18) gives

$$u \cdot (f(u) - f(v)) \geq 0$$

so f satisfies the constraint (u, v) .

This shows that S is sufficient.

It remains to prove that if $u^*, v^* \in T$ are T_w -opposed, and u^* is not equivalent to $-v^*$, then (u^*, v^*) is not provable from S by adding up. By Proposition 13, if (u^*, v^*) were provable from S by adding up, then it would be provable from $S_{[u^*, v^*)}$ by adding up. So it suffices to show that the latter is not the case.

For any $\alpha \in [0, 1]$, let $u_\alpha = (1 - \alpha)u^* + \alpha v^*$. Let $\alpha^* \in (0, 1)$ be such that $u_{\alpha^*} \in T_w$. Notice that if u, v are equivalent to u_α, u_β respectively, and $(u, v) \in S$, then α, β are either both less than α^* or both greater than α^* : otherwise u, v are T_w -opposed.

Suppose that (u^*, v^*) is provable from $S_{[u^*, v^*)}$ by adding up. Let c_{uv} be the coefficients that satisfy (11). Let $T_<$ be the set of types in $S_{[u^*, v^*)}$ that are equivalent to some u_α for $\alpha < \alpha^*$. The observation of the previous paragraph implies that if $c_{uv} > 0$, and one of u, v is in $T_<$, then the other is as well.

Sum up (11) over all $u \in T_<$. The $c_{uv}\bar{u}$ terms on the left side appear in pairs of opposite sign, which cancel; thus we are left with $0 = \bar{u}^*$. Since $u^* \in T$ cannot be indifferent, we have a contradiction. $\square$

4 Insufficiency results

The previous section offered numerous examples of type spaces on which local incentive constraints are sufficient. This naturally raises the question of whether there are important cases where local incentive constraints are insufficient. In this section we discuss such cases. First, we show that on some cardinal type spaces — roughly, those which violate convexity in a strong enough way — there are mechanisms that are locally incentive-compatible but not fully incentive-compatible, and moreover one can construct

such mechanisms so that they satisfy most of the possible incentive constraints. Second, we consider the particular case of dichotomous preferences, and show that no incomplete set of incentive constraints is sufficient to ensure full incentive-compatibility unless it is very close to being complete. After presenting these cases, we discuss the potential implications of insufficiency results (or lack thereof) for a mechanism designer.

4.1 Cardinal type spaces

In Subsection 3.1, we showed that on any convex cardinal type space, any set of local incentive constraints was sufficient. In this subsection, we show that if convexity is violated in a strong enough way, then this ceases to be true. (We do not have an exact characterization of the cardinal type spaces T for which every set of local incentive constraints is sufficient.)

In the space $\mathbb{R}^m$, let Π be the subspace of vectors whose sum of components is zero (as in Subsection 3.6). Let a *fair open half-space* be a set of the form $H = \{u \mid u \cdot \lambda > 0\}$ for some nonzero $\lambda \in \Pi$. Say that a cardinal type space T is *fairly separated* if there is some fair open half-space H such that the set $T \cap H$ is not connected.

Our main result here is:

Proposition 15 *Let T be a cardinal type space that is fairly separated. Then there exists a set of local incentive constraints that is not sufficient.*

Fair separatedness certainly implies nonconvexity: if T were convex, then for any fair open half-space H , $T \cap H$ would be convex and so connected; so T could not be fairly separated. Thus fair separatedness is at least logically a strengthening of nonconvexity. To further explain the relationship between the two concepts, a little graphical intuition is in order.

For concreteness, suppose X has four elements. Any utility function u is equivalent to its orthogonal projection onto Π ; this effectively reduces the set of possible utility functions to the three-dimensional space Π . Moreover, by rescaling by a constant, we can reduce every utility function to lie on the unit sphere in Π , except for the one at the origin. Thus we can think of the unit sphere, together with the origin, as representing all utility functions, up to equivalence. This sphere is illustrated in Figure 3. Then, the upper hemisphere, whose boundary is shown dashed in the figure, corresponds to a fair open half-space. If T contains the types labeled u and v , but does not contain any type along the thick curve (or any type equivalent to it), then T is fairly separated, hence there are local incentive constraints that are not sufficient.

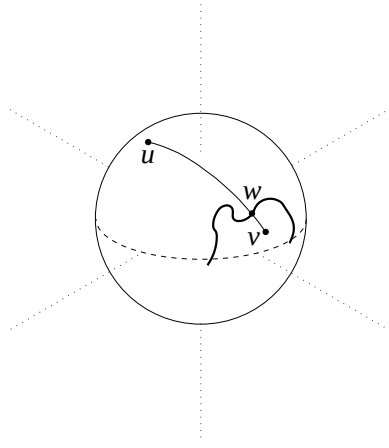


Figure 3: Illustration of fair separatedness

Notice that if we knew only that T contained u and v but not w (or any equivalent type), then T would be nonconvex. This is *not* enough to ensure that there are local incentive constraints that are insufficient — this is exactly the example in Proposition 14. On the other hand, if T excludes the whole curve, rather than just one point, then that is enough, by Proposition 15. This is why we think of fair separatedness as a strengthening of nonconvexity.¹

Fair separatedness might not seem like a natural condition on a type space, so we give one important example. We have already discussed single-peaked ordinal types, in Subsection 3.3. Consider now single-peaked cardinal types, and allow indifferences (if we do not allow indifferences then the type space is clearly disconnected). Then single-peakedness becomes equivalent to quasiconcavity. That is, let $x_1, \dots, x_m$ be the elements of X in a fixed order, and suppose $m \geq 3$. A utility function u on X is quasiconcave if there is some k such that whenever $i < j \leq k$ or $i > j \geq k$, we have $u(x_j) \geq u(x_i)$. Let T be the cardinal type space consisting of all quasiconcave utility functions.

Then T is fairly separated. For example, take any $1 \leq i < j < k \leq m$, and let $H = \{u \mid u(x_i) - 2u(x_j) + u(x_k) > 0\}$. If $u \in T \cap H$, then either $u(x_i) > u(x_j)$ or $u(x_k) > u(x_j)$. So $\{u \in T \cap H \mid u(x_i) > u(x_j)\}$ and $\{u \in T \cap H \mid u(x_k) > u(x_j)\}$ are two open, nonempty subsets of $T \cap H$, whose union is all of $T \cap H$, and whose intersection is

¹The reader might wonder whether four outcomes are really needed to illustrate fair separatedness. One can construct examples with two or three outcomes. But it would arguably be natural to restrict attention to type spaces T that are *regular* in the following sense: if $u \in T$, then any type equivalent to u is also in T . For regular type spaces, fair separatedness is equivalent to nonconvexity (aside from inclusion of indifferent types) unless there are at least four outcomes.

empty (any u satisfying both inequalities would violate quasiconcavity). Hence, $T \cap H$ is not connected.

Proof of Proposition 15: Let $H = \{u \mid u \cdot \lambda > 0\}$ with $\lambda \in \Pi$. Notice first that there are lotteries L, L' on X such that $H = \{u \mid u \cdot L > u \cdot L'\}$. Indeed, let L be any lottery with full support, and let $L' = L - \delta\lambda$, where $\delta > 0$ is chosen small enough so that all components of L' are still positive. Since $\lambda \in \Pi$, L' is again a lottery, and $u \cdot L > u \cdot L'$ if and only if $u \cdot \lambda > 0$.

Now write $T \cap H = T_a \cup T_b$, where T_a, T_b are open, disjoint, and nonempty. Consider the following mechanism:

- $f(u) = L$ if $u \in T_a$;
- $f(u) = L'$ otherwise.

Let S contain all incentive constraints (u, v) such that

- $u \in T \setminus T_b$ and $v \in T$; or
- $u \in T_b$ and $v \in T \setminus T_a$.

These are local incentive constraints. If $u \in T_a$, then T_a is a neighborhood of u and $(u, v), (v, u) \in S$ for all $v \in T_a$. If $u \in T_b$, then similarly for T_b . And otherwise, $u \in T \setminus H$, and then $(u, v), (v, u) \in S$ for any $v \in T$ whatsoever.

And f satisfies the incentive constraints S . Indeed, the image of f is $\{L, L'\}$. If $u \in T_a$ then $f(u) = L$, and if $u \in T \setminus H$, then $f(u) = L'$; either way, $f(u)$ is (weakly) preferred by u to the other possible value of f , so $u \cdot (f(u) - f(v)) \geq 0$ for any $v \in T$. If $u \in T_b$ and $v \in T \setminus T_a$, then $f(u) = f(v) = L'$ so that $u \cdot (f(u) - f(v)) = 0$.

Finally, f is not fully incentive-compatible: choose $u \in T_b$, $v \in T_a$, and then $f(u) = L'$, $f(v) = L$, and $u \cdot (f(u) - f(v)) < 0$. Thus, the incentive constraints S are not sufficient.

□

In the proof of this proposition, the set S constructed consists of all possible incentive constraints except those of the form (u, v) where $u \in T_b$ and $v \in T_a$. Depending on the situation, this may be a fairly large set. But it may not be large enough, and a mechanism designer interested in understanding locally incentive-compatible mechanisms might be concerned that an agent whose true utility function is in T_b would consider misreporting as some type in T_a . We indicate now how one can construct mechanisms that satisfy a larger set of incentive

constraints, but still are not fully incentive-compatible. The point is that the conclusion of Proposition 15 is fairly robust to the choice of local incentive constraints.

For any $\epsilon \geq 0$, let $B_\Pi(0, \epsilon)$ be the closed ball of radius ϵ about the origin in Π (under the Euclidean distance in $\mathbb{R}^m$). Similarly, for any lottery L , let $B_{\Delta(X)}(L, \epsilon)$ be the closed ball of radius ϵ about L in $\Delta(X)$.

For any utility function u , let $\hat{u}$ denote the point of $B_\Pi(0, 1)$ whose dot product with u is highest. (In the representation of Figure 3, $\hat{u}$ is the point on the sphere corresponding to u . $\hat{u}$ is uniquely defined, unless u is indifferent, in which case we let $\hat{u} = 0$.) For any lottery L and any ϵ , then, $L + \epsilon\hat{u}$ is the point of $B_{\Delta(X)}(L, \epsilon)$ preferred by the utility function u , as long as this point belongs to $\Delta(X)$.

Now, as in the preceding proof, write $T \cap H = T_a \cup T_b$, and let lotteries L, L' with full support satisfy $L' = L - \delta\lambda$. For $\epsilon \geq 0$, define the mechanism f on T as follows:

- if $u \in T_a$, then $f(u) = L + \epsilon\hat{u}$;
- otherwise, $f(u) = L' + \epsilon\hat{u}$.

As long as δ, ϵ are small enough, these values are actual lotteries (none of the probabilities specified are negative). When $\epsilon = 0$ we recover the mechanism defined in the proof above.

It is straightforward to check that this f still satisfies all the incentive constraints in $S = ((T \setminus T_b) \times T) \cup (T_b \times (T \setminus T_a))$, using the fact that $u \cdot (\hat{u} - \hat{v}) \geq 0$ for any u, v . So f is locally incentive-compatible, and in fact the only incentive constraints it can fail to satisfy are those of the form (u, v) with $u \in T_b, v \in T_a$. In this case, $f(u) - f(v) = -\delta\lambda + \epsilon(\hat{u} - \hat{v})$, so we see that $u \cdot (f(u) - f(v)) \geq 0$ is equivalent to

$$\frac{u \cdot \hat{u} - u \cdot \hat{v}}{u \cdot \lambda} \geq \frac{\delta}{\epsilon}. \quad (19)$$

Let $R(u, v)$ denote the left-hand side of (19). Then suppose the small positive numbers ϵ, δ are chosen so that δ/ϵ is appropriately small. Specifically, suppose δ/ϵ is close to, but larger than, $\inf_{u \in T_b, v \in T_a} R(u, v)$. Then most pairs (u, v) will satisfy (19). That is, f will satisfy most of the incentive constraints (u, v) , with $u \in T_b, v \in T_a$, and all incentive constraints not of this form, but it will still not be fully incentive-compatible.

4.2 Dichotomous preferences

If T is a *finite* cardinal type space (with more than one element), then it is certainly disconnected. In this case, local incentive constraints as we have defined them are uninteresting (the trivial constraints $\{(u, u) \mid u \in T\}$ qualify as local); on the other hand, the

above discussion shows us how to construct fairly large sets of incentive constraints that are not sufficient. In fact, however, we can get stronger results.

Suppose $u, v \in T$ are distinct types such that u is not indifferent nor equivalent to $-v$, and such that T contains no types equivalent to convex combinations of u and v except for u, v themselves. If S is any set of incentive constraints with $(u, v) \notin S$, then S cannot be sufficient. Indeed, for S to imply (u, v) , we know from Propositions 11 and 13 that (u, v) would have to be provable by adding up from $S_{[u,v]} \subseteq \{(u, u), (v, u), (v, v)\}$; one easily checks that this cannot happen. This argument shows that, on finite cardinal type spaces, a set S of incentive constraints typically needs to contain most of the possible constraints in $T \times T$ in order to be sufficient.

In the present subsection we further expand on this observation for a particular finite cardinal type space that has received much attention in mechanism design. The *type space of dichotomous preferences* (over a fixed outcome space X) is the cardinal type space consisting of all utility functions u such that $u(x) \in \{0, 1\}$ for all $x \in X$. For an agent with such a utility function, each outcome is either acceptable or unacceptable; there are no further grades of preference. Dichotomous preferences may arise, for example, for workers scheduling a meeting (for each possible time, any given worker is either able or unable to attend); they also arise in organ transplantation (each donor is compatible or incompatible with a given recipient). Bogomolnaia, Moulin, and Stong [15] investigate the general social choice problem with dichotomous preferences; matching mechanisms for dichotomous preferences are studied by Bogomolnaia and Moulin [14] and Roth, Sönmez, and Ünver [39].

Continuing to abuse notation, we can identify types in this space with subsets of X : the utility function u is identified with the set $\{x \mid u(x) = 1\}$. A mechanism then maps subsets of X into lotteries over X ; an incentive constraint is an ordered pair of subsets of X .

There are several situations where an incentive constraint can be implied by a combination of other constraints. For example, the incentive constraints $(\emptyset, Y)$ and (X, Y) are always satisfied by any mechanism (because the utility functions corresponding to $\emptyset$ and X are both indifferent among all outcomes). For a less trivial example, suppose $Y, Z \subseteq X$ with $Y \cap Z = \emptyset$. Then, if a mechanism satisfies the incentive constraints $(Y, Y \cup Z)$, $(Z, Y \cup Z)$, and $(Y \cup Z, Z)$, it must also satisfy (Y, Z) . This is easily shown by the usual adding-up argument from the proof of Proposition 1, noting that the utility function $Y \cup Z$ is equivalent to $(Y + Z)/2$. Similarly, if $Y \cup Z = X$, then any mechanism that satisfies $(Y, Y \cap Z)$, $(Z, Y \cap Z)$, and $(Y \cap Z, Z)$ must also satisfy (Y, Z) .

However, these are exceptional cases: most incentive constraints are not implied by any combination of other incentive constraints. This is immediate from the discussion at the beginning of this subsection, but let us give a full, explicit statement.

Proposition 16 *Suppose that T is the type space of dichotomous preferences over X , and S is a set of incentive constraints. Suppose $Y, Z \subseteq X$ with $(Y, Z) \notin S$, and suppose that $Y \neq Z$, $Y \cap Z \neq \emptyset$ and $Y \cup Z \neq X$. Then, there exists a mechanism f satisfying S but not satisfying the incentive constraint (Y, Z) . In particular, S is not sufficient.*

Proof: If S implies (Y, Z) , then by Proposition 11, (Y, Z) is provable from S by adding up. The hypotheses ensure that Proposition 13 applies, so in fact (Y, Z) is provable from $S_{[Y, Z]}$ by adding up. However, there is no type in T equivalent to $(1 - \alpha)Y + \alpha Z$ for any $\alpha \in (0, 1)$, since any such convex combination is a utility function with at least three distinct values (namely 0 on any outcome in $X \setminus (Y \cup Z)$, 1 on any outcome in $Y \cap Z$, and α or $1 - \alpha$ on any outcome in $Y \setminus Z$ or $Z \setminus Y$), but utility functions in T can only have two distinct values. So the only types in $T_{[Y, Z]}$ must be equivalent to Y or Z . Since Y and Z are not equal to $\emptyset$ or X , the only such types are Y, Z themselves. Hence $S_{[Y, Z]} \subseteq \{(Y, Y), (Z, Z), (Z, Y)\}$. It is easy to check that (Y, Z) is not provable from such a set by adding up. So we have a contradiction, and in fact S cannot imply (Y, Z) . $\square$

When X is large, almost all of the possible incentive constraints (Y, Z) satisfy the conditions $Y \neq Z, Y \cap Z \neq \emptyset, Y \cup Z \neq X$ of the proposition. More specifically, of the 4^m incentive constraints one could write down (one for each of the 2^m possible choices of Y and 2^m possible choices of Z), only $2 \cdot 3^m - 2$ violate the conditions, and $(2 \cdot 3^m - 2)/4^m \rightarrow 0$ as $m \rightarrow \infty$. Thus, Proposition 16 implies that, in order to be sufficient, the set S must contain most of the possible incentive constraints.

At this point we turn to considering whether our conclusions apply to deterministic mechanisms. In Section 3, there was no need to consider deterministic mechanisms separately: the sufficiency results held for all probabilistic mechanisms, so they automatically applied to deterministic mechanisms. With our insufficiency results, however, it may be that some set of constraints is not sufficient for probabilistic mechanisms but still sufficient for deterministic mechanisms. (We raise the issue only now, not in Subsection 4.1, because the present type spaces are specific enough to give concrete results.)

In our setting of dichotomous preferences, a deterministic mechanism is one for which $f(Y)$ is a degenerate lottery for all $Y \subseteq X$. Taking a notational shortcut once again, we now think of $f(Y)$ as an element of X . We call a set of incentive constraints S *sufficient for deterministic mechanisms* if every deterministic mechanism that satisfies S

is fully incentive-compatible. Notice that an incentive constraint (Y, Y') is satisfied unless $f(Y) \notin Y$ and $f(Y') \in Y$.

Our first observation on deterministic mechanisms is that the lesson of Proposition 16 does not apply: there are sets of incentive constraints that are sufficient for deterministic mechanisms and are relatively sparse. That is, there exists a sufficient set S_m of incentive constraints when $|X| = m$, such that $|S_m|/4^m \rightarrow 0$ as $m \rightarrow \infty$.

Proposition 17 *Suppose that T is the type space of dichotomous preferences over X . Let S be a set of incentive constraints such that*

- *for every $Y \subseteq X$ with $|Y| \geq 2$, and every $x \in Y$, there exists a proper subset Y' of Y , containing x , such that $(Y, Y') \in S$;*
- *for every $Y \subseteq X$ with $|Y| = 1$, and every $Y' \subseteq X$, $(Y, Y') \in S$.*

Then S is sufficient for deterministic mechanisms.

Proof: Let f be a deterministic mechanism satisfying the incentive constraints S . For all Y , either $f(Y) \in Y$, or $f(Y) \notin Y$. Call Y *pleased* in the former case and *displeased* in the latter. Suppose that f fails to satisfy some incentive constraint $(Z, Z') \notin S$. So Z is displeased, whereas $f(Z') \in Z$.

Now let Y be a minimal displeased set containing $f(Z')$. (Some such displeased set exists, namely Z .) If $|Y| \geq 2$, then by hypothesis there is some proper subset $Y' \subset Y$ with $f(Z') \in Y'$ such that $(Y, Y') \in S$. By minimality, Y' is pleased: $f(Y') \in Y' \subset Y$. But since $f(Y) \notin Y$, then, f violates the incentive constraint (Y, Y') , a contradiction.

If on the other hand $|Y| = 1$, then $Y = \{f(Z')\}$. Now $f(Y) \notin Y$ (because Y is displeased) but $f(Z') \in Y$, so f violates the incentive constraint $(Y, Z') \in S$, again a contradiction. $\square$

On the other hand, the set of incentive constraints cannot be made too sparse without losing sufficiency:

Proposition 18 *Suppose that T is the type space of dichotomous preferences over X . Let $x \in X$, and let Z be any set containing x and not equal to $\{x\}$. If S is a set of incentive constraints such that $(\{x\}, Z) \notin S$, then there exists a deterministic mechanism satisfying S but not satisfying the constraint $(\{x\}, Z)$. In particular, S is not sufficient for deterministic mechanisms.*

This proposition in particular implies that, for deterministic mechanisms over dichotomous type spaces, the trial-and-error manipulation argument for computationally limited

agents (as sketched out in the introduction) does not apply. Suppose we are designing a mechanism for an agent, and we predict that if the agent's true type is $\{x\}$, then he will contemplate only misreports in some known set, which may vary with m but whose size grows more slowly than 2^m . Then there are deterministic mechanisms that are not fully incentive-compatible but that the agent will nonetheless fail to manipulate, for large enough m .

Proof: We give a construction. Define f as follows:

- $f(\{x\})$ is any element of X other than x ;
- $f(Z) = x$;
- $f(\emptyset)$ is any element of X other than x ;
- for any other Y , $f(Y)$ is an arbitrary element of Y distinct from x .

Using the terminology from the proof of Proposition 17, every set is pleased except for $\{x\}$ and $\emptyset$. Therefore, f satisfies every possible incentive constraint (Y, Y') except possibly when $Y = \{x\}$ or $Y = \emptyset$. If $Y = \emptyset$, then $f(Y') \notin Y$ for all Y' , so (Y, Y') is again satisfied. Finally, if $Y = \{x\}$, then by construction $f(Y') \notin Y$ unless $Y' = Z$. So f fails to satisfy the incentive constraint $(\{x\}, Z)$, but it satisfies every other possible incentive constraint. $\square$

We can reach the following conclusions for deterministic mechanisms on dichotomous type spaces. The strong insufficiency result for probabilistic mechanisms (Proposition 16) no longer holds; a set S of incentive constraints can be sufficient for full incentive-compatibility without needing to contain most of the incentive constraints in $T \times T$. On the other hand, if we specify that every single agent type can only consider a small number of misreports, i.e. for each Y there are a small number of Y' such that $(Y, Y') \in S$, then we once again get that S is not sufficient.

4.3 Implications of insufficiency

To put our insufficiency results more fully into context, we now return to our main motivating question: Can one get around impossibility results such as the Gibbard-Satterthwaite theorem by substituting local incentive-compatibility conditions in place of full incentive-compatibility? For type spaces where local incentive constraints are sufficient, the answer is definitively no. However, when local incentive constraints are insufficient, the answer may or may not be yes.

We illustrate with a simple example, voting rules. For simplicity we restrict attention to deterministic mechanisms in this discussion. Suppose that X consists of $m \geq 4$ outcomes, and T is the full ordinal type space. Suppose there are $n \geq 2$ agents. A *voting rule* is a function $f : T^n \rightarrow X$, specifying a single outcome for each profile of the agents' preferences. As in the discussion at the end of Subsection 2.1, we say that a voting rule *satisfies* a set S of incentive constraints if, for each agent i , each constraint $(t_i, t'_i) \in S$ and each profile t_{-i} of types of the other agents, t_i weakly prefers $f(t_i, t_{-i})$ over $f(t'_i, t_{-i})$. We also say that the voting rule is *anonymous* if, for all types $t_1, \dots, t_n \in T$ and every permutation π of the agents $1, \dots, n$, we have $f(t_1, \dots, t_n) = f(t_{\pi(1)}, \dots, t_{\pi(n)})$. The voting rule is *dictatorial* if there exists some fixed agent i such that $f(t_1, \dots, t_n)$ equals the outcome ranked highest by t_i , for all $t_1, \dots, t_n$. It is *unanimous* if, whenever all types $t_1, \dots, t_n$ have the same top-ranked outcome x , then $f(t_1, \dots, t_n) = x$.

Let $S = T \times T$ be the full set of incentive constraints. Then one statement of the Gibbard-Satterthwaite theorem [22, 43] in this context is as follows:

Proposition 19 *Let f be a unanimous voting rule that satisfies S . Then f is dictatorial.*

A slightly weaker statement is:

Proposition 20 *Let f be a unanimous voting rule that satisfies S . Then f is not anonymous.*

What happens when we replace S by some insufficient set of incentive constraints? Consider the set S' consisting of all constraints (t, t') such that t and t' have the same bottom-ranked outcome. This set is insufficient for deterministic mechanisms (for example, it is satisfied by the one-agent mechanism that assigns each type its bottom-ranked outcome). If we replace S by S' , then Proposition 19 no longer holds, but Proposition 20 does still hold.

To see that Proposition 19 is overturned, just consider any voting rule that chooses a different dictator depending on the agents' bottom choices. For example, for some fixed $x \in X$, let $f(t_1, \dots, t_n)$ be t_1 's top-ranked outcome if x is bottom-ranked by t_1 , and otherwise let it be t_2 's top-ranked outcome. This is unanimous, satisfies S' (because each agent cannot change the identity of the dictator without changing her bottom-ranked outcome), and is not dictatorial.

To see that Proposition 20 still holds, just fix any $x \in X$ and restrict attention to the type profiles at which x is every agent's bottom-ranked outcome.

More precisely: Let f be unanimous and satisfy S' . If all agents rank x bottom, then f cannot choose the outcome x . To see this, suppose there does exist some profile where all

agents rank x at the bottom, yet f chooses x . Fix an outcome $y \in X$ distinct from x , and consider such a profile for which the number of agents ranking y at the top is maximal. There must be some agent i whose type t_i does not rank y at the top, since otherwise f would choose y , by unanimity. But then, if i instead reported a type t'_i ranking y at the top and x at the bottom, the outcome would not be x (by maximality). This means that f violates the incentive constraint $(t_i, t'_i) \in S'$, a contradiction.

So, let $\tilde{X} = X \setminus \{x\}$ and $\tilde{T}$ be the full ordinal type space on $\tilde{X}$, and let $\tilde{S} = \tilde{T} \times \tilde{T}$. We can define a voting rule $\tilde{f} : \tilde{T}^n \rightarrow \tilde{X}$, simply by appending x to the bottom of each agent's preference ranking and applying f . The previous paragraph ensures that the image of $\tilde{f}$ is contained in $\tilde{X}$. Because f is unanimous, so is $\tilde{f}$; because f satisfies S' , $\tilde{f}$ satisfies $\tilde{S}$. So by the Gibbard-Satterthwaite theorem, $\tilde{f}$ is dictatorial, and in particular it cannot be anonymous. This implies that f cannot be anonymous either.

The point of this discussion is that, by replacing full incentive-compatibility with satisfaction of some insufficient set of incentive constraints, we may or may not succeed in overturning impossibility theorems.

Admittedly the example does not perfectly match the modeling approach we have taken, since S' is not a set of local incentive constraints. A slightly more complicated variant is immune to this criticism. Suppose $m \geq \max\{5, n+2\}$, and fix $y \in X$. Let T be the set of all ordinal types for which y is ranked second-to-last. Then arguments similar to the above again show that the analogues of Propositions 19 and 20 both hold when $S = T \times T$; if instead we use the set S' of local incentive constraints, Proposition 19 no longer holds, while Proposition 20 still does.

Even if it is possible to overturn impossibility theorems by requiring only local incentive-compatibility, a mechanism designer who accepts this relaxation should be aware that doing so involves strong implicit assumptions about the limits to agents' sophistication. For example, simply believing that agents are computationally limited is not enough justification. Suppose that the designer works with a set S of local incentive constraints in some type space that is not sufficient, but there is some other set S' of incentive constraints that is sufficient, and for every t the set $\{t' \mid (t, t') \in S'\}$ is small. Then agents with limited computational power but adequate forethought might simply decide to use S' rather than S when looking for advantageous misreports, and thereby end up finding some. Thus, in being satisfied with mechanisms that satisfy the constraints S , the designer must be fairly committed to the assumption that agents will not look elsewhere for potential manipulations.

Alternatively, suppose that the mechanism designer wants to use local incentive-

compatibility to overturn some known impossibility theorem: no mechanism exists satisfying full incentive-compatibility (together with some other desiderata), but the mechanism designer finds an f that satisfies only local incentive-compatibility instead. Suppose there is a published proof of the theorem available that is “explicit” — that is, it proceeds by contradiction, supposing the desired fully incentive-compatible mechanism exists, and each time incentive-compatibility is applied, the misreport under consideration is fully spelled out. If the agents are aware of this proof, they can simply go through it step by step and eventually encounter the step which fails to hold for the given f ; this step must use some nonlocal misreport that is potentially advantageous under f . Thus, agents who are aware enough to read an explicit proof may discover a profitable misreport, contrary to the designer’s expectation. (To this author’s knowledge, every known impossibility result has an explicit proof available in the literature.)

All of these issues are somewhat unimportant in the grand scheme of things, since on the type spaces for which we have shown insufficiency results (in particular, the dichotomous type space), the known mechanism design results are mostly positive ones. The point is simply that one should be cautious in interpreting insufficiency results. We have presented such results in order to show that the ideas of Section 3 do not apply everywhere; but the results by themselves do not argue for or against imposing local rather than full incentive-compatibility constraints. If a practitioner wants to impose only the requirement of local incentive-compatibility, there had better be reasonable a priori grounds for doing so.

5 Conclusion

This paper has examined the question of whether or not a small set of incentive constraints is sufficient to ensure that all other incentive constraints are automatically satisfied, in many common mechanism design settings. In particular, we have given an arguably natural definition of “local” incentive constraints for cardinal and polyhedral (including ordinal) type spaces. For many type spaces, we have shown that local incentive constraints are sufficient to imply full incentive-compatibility. This holds for convex spaces of cardinal types; polyhedral type spaces with convex closure; type spaces of single-peaked ordinal preferences (as well as single-troughed ordinal preferences, or single-peaked ordinal preferences on a tree); and for type spaces of successive single-crossing ordinal preferences. Additionally, while we have focused on mechanisms that specify distributions over finite outcome spaces, we also showed that local incentive constraints are sufficient over convex

cardinal type spaces in settings where monetary transfers are possible and utilities are quasilinear. All of these proofs consisted of obtaining any desired incentive constraint as a sum of inequalities associated with local incentive constraints; we also gave an example where local incentive constraints ensure full incentive-compatibility but no such proof by adding up exists.

These results provide an immediate strengthening of many existing impossibility and characterization theorems in mechanism design, as well as a possible shortcut to proving future incentive-compatibility results.

On the other hand, for cardinal type spaces meeting a strong nonconvexity condition, we have insufficiency — there are probabilistic mechanisms that are locally incentive-compatible but not fully incentive-compatible. This result is robust to the choice of local incentive constraints, in the sense that there are mechanisms that satisfy most incentive constraints but still are not fully incentive-compatible. In particular, for the dichotomous type space over a large outcome space, almost every incentive constraint has the property that it can be violated by a mechanism that nonetheless satisfies all other incentive constraints. However, our results should not immediately be interpreted as a license to impose only local incentive-compatibility rather than full incentive-compatibility in these type spaces.

Our work here is in response to a broader conceptual challenge in mechanism design: how to formulate incentive constraints that are strong enough to ensure that agents' behavior in the resulting mechanisms can be accurately predicted, without being overly restrictive. The usual notion of incentive-compatibility says that no type should have an incentive to imitate any other type; but in practice, this requirement may be unnecessarily strong, if the task of finding a profitable misreport is difficult for the agents. Thus, thinking about mechanisms that are hard, but not impossible, to manipulate may provide a useful avenue for getting around negative results. This paper has explored one version of this approach, by imagining that agents of any given type consider only a limited subset of the possible misreports, focusing on misreports that are close to the true type. Our results show many settings in which this change does not affect the set of possible mechanisms, and some in which it does.

References

- [1] Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth (2005), “The New York City High School Match,” *American Economic Review* 95 (2), 364-367.

- [2] José Alcalde and Salvador Barberà (1994), “Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems,” *Economic Theory* 4 (3), 417-435.
- [3] Aaron Archer and Robert Kleinberg (2008), “Truthful Germs are Contagious: A Local to Global Characterization of Truthfulness,” in *Proceedings of the 9th ACM Conference on Electronic Commerce* (EC-08), 21-30.
- [4] Itai Ashlagi, Mark Braverman, Avinatan Hassidim, and Dov Monderer (forthcoming), “Monotonicity and Implementability,” *Econometrica*.
- [5] Salvador Barberà, Anna Bogomolnaia, and Hans van der Stel (1998), “Strategy-Proof Probabilistic Rules for Expected Utility Maximizers,” *Mathematical Social Sciences* 35 (2), 89-103.
- [6] Salvador Barberà, Matthew O. Jackson, and Alejandro Neme (1997), “Strategy-Proof Allotment Rules,” *Games and Economic Behavior* 18 (1), 1-21.
- [7] Salvador Barberà, Hugo Sonnenschein, and Lin Zhou (1991), “Voting by Committees,” *Econometrica* 59 (3), 595-609.
- [8] John J. Bartholdi III and James B. Orlin (1991), “Single Transferable Vote Resists Strategic Voting,” *Social Choice and Welfare* 8 (4), 341-354.
- [9] J. J. Bartholdi III, C. A. Tovey, and M. A. Trick (1989), “The Computational Difficulty of Manipulating an Election,” *Social Choice and Welfare* 6 (3), 227-241.
- [10] Sushil Bikhchandani, Shurojit Chatterji, Ron Lavi, Ahuva Mu’alem, Noam Nisan, and Arunava Sen (2006), “Weak Monotonicity Characterizes Deterministic Dominant-Strategy Implementation,” *Econometrica* 74 (4), 1109-1132.
- [11] Duncan Black (1948), “On the Rationale of Group Decision-Making,” *Journal of Political Economy* 56 (1), 23-34.
- [12] Anna Bogomolnaia and Hervé Moulin (2001), “A New Solution to the Random Assignment Problem,” *Journal of Economic Theory* 100 (2), 295-328.
- [13] Anna Bogomolnaia and Hervé Moulin (2002), “A Simple Random Assignment Problem with a Unique Solution,” *Economic Theory* 19 (3), 623-635.

- [14] Anna Bogomolnaia and Hervé Moulin (2004), “Random Matching under Dichotomous Preferences,” *Econometrica* 72 (1), 257-279.
- [15] Anna Bogomolnaia, Hervé Moulin, and Richard Stong (2005), “Collective Choice under Dichotomous Preferences,” *Journal of Economic Theory* 122 (2), 165-184.
- [16] Gorkem Celik (2006), “Mechanism Design with Weaker Incentive Compatibility Constraints,” *Games and Economic Behavior* 56 (1), 37-44.
- [17] Vincent Conitzer and Tuomas Sandholm (2006), “Nonexistence of Voting Rules That Are Usually Hard to Manipulate,” in *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, 627-634.
- [18] Vladimir I. Danilov (1994), “The Structure of Non-Manipulable Social Choice Rules on a Tree,” *Mathematical Social Sciences* 27 (2), 123-131.
- [19] Gabrielle Demange (1982), “Single-Peaked Orders on a Tree,” *Mathematical Social Sciences* 3 (4), 389-396.
- [20] Catherine Eckel and Charles A. Holt (1989), “Strategic Voting in Agenda-Controlled Committee Experiments,” *American Economic Review* 79 (4), 763-773.
- [21] Lars Ehlers, Hans Peters, and Ton Storcken (2002), “Strategy-Proof Probabilistic Decision Schemes for One-Dimensional Single-Peaked Preferences,” *Journal of Economic Theory* 105 (2), 408-434.
- [22] Allan Gibbard (1973), “Manipulation of Voting Schemes: A General Result,” *Econometrica* 41 (4), 587-601.
- [23] Allan Gibbard (1977), “Manipulation of Schemes that Mix Voting with Chance,” *Econometrica* 45 (3), 665-681.
- [24] Allan Gibbard (1978), “Straightforwardness of Game Forms with Lotteries as Outcomes,” *Econometrica* 46 (3), 595-614.
- [25] Jerry R. Green and Jean-Jacques Laffont (1986), “Partially Verifiable Information and Mechanism Design,” *Review of Economic Studies* 53 (3), 447-456.
- [26] Glenn W. Harrison and Tanga McDaniel (2008), “Voting Games and Computational Complexity,” *Oxford Economic Papers* 60 (3), 546-565.

- [27] Roberta Q. Herzberg and Rick K. Wilson (1988), “Results on Sophisticated Voting in an Experimental Setting,” *Journal of Politics* 50 (2), 471-486.
- [28] Aanund Hylland (1980), “Strategy Proofness of Voting Procedures with Lotteries as Outcomes and Infinite Sets of Strategies,” unpublished paper, Harvard University.
- [29] Marcus Isaksson, Guy Kindler, and Elchanan Mossel (2010), “The Geometry of Manipulation — a Quantitative Proof of the Gibbard Satterthwaite Theorem,” ArXiv preprint, <http://arxiv.org/abs/0911.0517> .
- [30] Bettina Klaus, Hans Peters, and Ton Storcken (1997), “Strategy-Proof Division of a Private Good when Preferences are Single-Dipped,” *Economics Letters* 55 (3), 339-346.
- [31] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green (1995), *Microeconomic Theory* (New York: Oxford University Press).
- [32] Eric Maskin and John Riley (1984), “Monopoly with Incomplete Information,” *RAND Journal of Economics* 15 (2), 171-196.
- [33] Allan H. Meltzer and Scott F. Richard (1981), “A Rational Theory of the Size of Government,” *Journal of Political Economy* 89 (5), 914-927.
- [34] Hervé Moulin (1980), “On Strategy-Proofness and Single Peakedness,” *Public Choice* 35 (4), 437-455.
- [35] Roger B. Myerson (1981), “Optimal Auction Design,” *Mathematics of Operations Research* 6 (1), 58-73.
- [36] Kevin W. S. Roberts (1977), “Voting over Income Tax Schedules,” *Journal of Public Economics* 8 (3), 329-340.
- [37] Jean-Charles Rochet (1987), “A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context,” *Journal of Mathematical Economics* 16 (2), 191-200.
- [38] Alvin E. Roth (1982), “The Economics of Matching: Stability and Incentives,” *Mathematics of Operations Research* 7 (4), 617-628.
- [39] Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver (2005), “Pairwise Kidney Exchange,” *Journal of Economic Theory* 125 (2), 151-188.

- [40] Michael Saks and Lan Yu (2005), “Weak Monotonicity Suffices for Truthfulness on Convex Domains,” in *Proceedings of the 6th ACM Conference on Electronic Commerce* (EC-05), 286-293.
- [41] Alejandro Saporiti (2009), “Strategy-Proofness and Single-Crossing,” *Theoretical Economics* 4 (2), 127-163.
- [42] Shin Sato (2009), “Adjacent Manipulation: A New Criterion on Manipulability of Social Choice Functions,” CAES Working Paper Series (Fukuoka University), #2009-015.
- [43] Mark A. Satterthwaite (1975), “Strategy-proofness and Arrow’s Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions,” *Journal of Economic Theory* 10 (2), 187-217.
- [44] Itai Sher and Rakesh Vohra (2010), “Optimal Selling Mechanisms on Incentive Graphs,” unpublished paper, University of Minnesota.
- [45] Yves Sprumont (1991), “The Division Problem with Single-Peaked Preferences: A Characterization of the Uniform Allocation Rule,” *Econometrica* 59 (2), 509-519.
- [46] Lin Zhou (1990), “On a Conjecture by Gale about One-Sided Matching Problems,” *Journal of Economic Theory* 52 (1), 123-135.