

Nonbossy Mechanisms: Mechanism Design with Almost-Vanishing Externalities

RENATO PAES LEME and JON SCHNEIDER, Google Research
HANRUI ZHANG, Carnegie Mellon university

We study mechanism design when agents have a non-zero but infinitesimally small interest in the utilities of each other. We show that in such cases, a mechanism is robust against strategic manipulation if and only if it is not only incentive-compatible, but also nonbossy — a well-studied property in the context of matching and allocation mechanisms. We give complete characterizations of incentive-compatible and nonbossy mechanisms in various settings, including auctions with single-parameter agents and public decision settings where all agents share a common outcome. In particular, we show that in the single-item setting, a mechanism is incentive-compatible, individually rational, and nonbossy if and only if it is a sequential posted-price mechanism. In contrast, we show that in more general single-parameter environments, there exist mechanisms satisfying our characterization that significantly outperform sequential posted-price mechanisms in terms of revenue or efficiency (sometimes by an exponential factor).

1 INTRODUCTION

Incentive-compatibility is the canonical notion of robustness against strategic behavior in mechanism design. Incentive-compatible mechanisms guarantee that *self-interested* agents cannot benefit from misreporting their private information. However, such mechanisms may still be prone to manipulation by agents who *care about each other*. For example, two rival companies competing for an item at an auction may not care solely about the utility they gain from winning the item, but may also care about the utility their rival loses from having to pay a higher price. In other settings, altruistic agents may not care solely about their own utility but also care some amount for the well-being of others. Such “externalities” complicate the standard picture of mechanism design.

Interestingly, many common incentive-compatible mechanisms are no longer robust against strategic behavior in the presence of these externalities, *even when these externalities are infinitesimally small*. For example, consider two weakly cooperating agents participating in a second price auction, where a single item is being sold. These agents behave in the following way: their primary goal is to maximize their own utility, which is their private value minus the payment if they win the item, and 0 otherwise; in the case where they cannot (rationally) win the item, they instead try to achieve their secondary goal, which is to maximize the other agent’s utility. Then, assuming both agents know each other’s value, one dominant bidding strategy for both agents is the following: the agent with the higher value should bid their true value, and the agent with the lower value should bid 0. As a result, the agent with the higher value still wins the item, but their payment is always 0.

On the other hand, there do exist some incentive-compatible mechanisms which remain robust in the presence of vanishingly small (but non-zero) externalities. One reason why sequential posted-price mechanisms — where an auctioneer iterates through the agents in order, offering each agent a take-it-or-leave-it price — are appealing in practice is this type of robustness. In particular, in a posted-price mechanism, an agent cannot change the utility of other agents without changing their own outcome (winning or losing the item), and thus cannot optimize for their secondary objective without costing themselves utility. Note that posted-price mechanisms are further agnostic to the specific types of externalities of the agents: they are robust regardless of whether the agents’ externalities are altruistic or malicious (or completely arbitrary), and do not require the agents to report their externalities up front.

The above examples illustrate the setting that we call “almost-vanishing externalities”, where externalities are present and unreported to the mechanism designer, but are less important than an agent’s primary goal of maximizing their own utility. Many rational agents in the real world naturally behave this way: such agents act carefully to optimize for their main interest, *except when it does not really matter* — if multiple actions are equally good as far as their main interest is concerned, then they turn to further optimizing for their interest in other agents, be it sympathy, envy, or any other consideration deemed rationally irrelevant.

In this paper, we study the problem of mechanism design in the setting of almost-vanishing externalities. Our main goal is to answer the following questions:

What mechanisms are robust against strategic behavior when agents exhibit almost-vanishing interest in each other’s utilities? What are the power and limitations of these mechanisms?

1.1 Our Results

Alternate interpretation: nonbossy mechanisms. We begin by giving another interpretation of robustness in the presence of almost-vanishing externalities: such robustness requires, in addition to the classical notion of incentive-compatibility, that no agent can change the outcomes that other agents receive, or the amounts that other agents pay, without changing their own outcome or

payment. We call mechanisms that satisfy this guarantee *nonbossy mechanisms* (in accordance with the existing notion of nonbossiness in the economics literature; see Section 1.2 for an overview). Moreover, incentive-compatibility and nonbossiness together are also sufficient for robustness in the presence of almost-vanishing externalities. This interpretation of externality-robust mechanisms as nonbossy mechanisms is more amenable to analysis and enables all subsequent results of the paper.

Characterizing nonbossy mechanisms. We then turn to the structure of nonbossy mechanisms in various concrete classes of environments that are commonly considered in the context of mechanism design. We first prove a general characterization of the payment rules of nonbossy mechanisms. Specifically, we show that as long as the environment is “expressive” enough (contains a wide enough range of valuation functions) the payment rule of any incentive-compatible and nonbossy mechanism is a function of the outcomes *only* – if two valuation profiles lead to the same outcomes for all agents, then the amounts that all agents pay must also be the same. Almost all commonly studied environments are expressive enough for this characterization to hold¹.

Based on the payment characterization, we further investigate two prominent settings in mechanism design, and provide complete characterizations of nonbossy mechanisms. First, we consider the abstract setting where the mechanism selects a single outcome (corresponding to a public decision, such as the location of a hospital) that is shared by all agents, and each agent may value each potential outcome arbitrarily. For this setting, we show that quite surprisingly, any mechanism that is incentive-compatible and nonbossy must either be dictatorial or always choose between only 2 outcomes. In other words, when choosing a common outcome, only trivial mechanisms can be robust against strategic behavior if agents care (even infinitesimally) about each other. This is reminiscent of the celebrated Gibbard-Satterthwaite theorem [Gibbard, 1973, Satterthwaite, 1975], and indeed the proof works by reducing to that theorem.

We then investigate the single-parameter setting where identical items are allocated, and each agent is interested in at most one item. For this setting, we show that nonbossy (and incentive-compatible and individually rational) mechanisms generalize sequential posted-price mechanisms, which approach all agents one by one and make take-it-or-leave-it offers. In particular, if there is a single item being sold, then the two classes of mechanisms are exactly the same. This provides an interesting semantic interpretation of sequential posted-price mechanisms in the single-item setting: the only mechanisms in the single-item setting that are robust in the presence of almost-vanishing externalities are sequential posted-price mechanisms. To the best of our knowledge, this is the first characterization of this kind for sequential posted-price mechanisms.

When there are multiple items (or even richer feasibility constraints), nonbossy, incentive-compatible and individually rational mechanisms are strictly richer than sequential posted-price mechanisms: in particular, they allow for posting a *vector* of prices to all agents simultaneously, which is accepted if and only if each agent accepts their personal component of the price vector. More generally, nonbossy, incentive-compatible and individually rational mechanisms correspond to *decision lists with exceptions*: a mechanism can be implemented using a sequence of price vectors, each associated with a distinct vector of outcomes and an exception list, consisting of a subset of other price vectors. The mechanism checks these price vectors one by one, and chooses the outcomes associated with the first vector satisfying: (1) all agents accept the price vector, and (2) there does not exist another price vector that is also accepted in the exception list of this price vector. In fact, we show that if the exception lists are chosen appropriately, then the order in which these price vectors are considered does not matter. The proof explicitly constructs a decision list given a mechanism, by first showing that any nonbossy and incentive-compatible mechanism must

¹We also show this property is necessary for this characterization, in the sense that there exist environments without this property in which the payment characterization fails.

“break ties consistently”, i.e., if the mechanism chooses one outcome vector when another outcome vector is also satisfied, then it can never choose the latter vector whenever the former is satisfied. Given this, one can construct a decision tree from a mechanism, which can then be turned into a decision list by merging nodes corresponding to the same outcome vector.

Separating nonbossy and sequential posted-price mechanisms. Our characterization shows that in single-parameter environments, nonbossy mechanisms generally have richer *forms* than sequential posted-price mechanisms. However, it is not immediately clear how much additional *power* nonbossy mechanisms have. To this end, we further establish strict separations in terms of revenue and efficiency between mechanisms that are nonbossy, incentive-compatible and individually rational, and sequential posted-price mechanisms in single-parameter environments. In particular, we identify two sources of separation: rich feasibility constraints and correlation between agents. With arbitrary downward-closed feasibility constraints, we show that nonbossy, incentive-compatible and individually rational mechanisms can guarantee an $\Omega(1/\log r)$ fraction of the optimal welfare (or revenue), where r is the maximum number of agents who can simultaneously receive an item. In contrast, it is known that no sequential posted-price mechanism can achieve an approximation ratio better than $O(1/r)$ [Babaioff et al., 2007]. When agents’ valuations can be correlated, we provide examples with only 2 agents where nonbossy, incentive-compatible and individually rational mechanisms can extract the full optimal welfare as revenue, whereas sequential posted-price mechanisms can only partially extract the optimal welfare. These results shed further light on the potential of nonbossy mechanisms, and show how they can provide more desirable tradeoffs between robustness against strategic behavior and revenue/efficiency compared to other simple mechanisms. In particular, nonbossy mechanisms are more powerful than sequential posted-price mechanisms (the archetypal “simple mechanisms”), and at the same time provide stronger robustness compared to generic incentive-compatible mechanisms.

1.2 Related Work

Nonbossy allocation rules and social choice functions. Our results can be viewed as a monetary version of previous results on nonbossy allocation rules and social choice functions. Satterthwaite and Sonnenschein [1981] propose the notion of nonbossiness for allocation mechanisms without money, and give several characterizations thereof. Since then, nonbossy allocation rules have been studied extensively: see the survey by Thomson [2016] for a comprehensive exposition. The vast majority of these results focus on environments without money (e.g., allocation of indivisible items [Hatfield, 2009, Pápai, 2001, Svensson, 1999], matching [Bade, 2020, Kojima, 2010], etc.), which make them incomparable with our results. Below we discuss several results that are particularly related to ours.

Svensson and Larsson [2002] study nonbossy allocation mechanisms where, in addition to indivisible items, there is a fixed amount of money being allocated. They show that the space of strategyproof and nonbossy mechanisms is finite, and that this space can be further restricted with additional properties enforced. This setting is different from ours, because the total “payment” (i.e., money being allocated) has to sum up to the fixed amount — indeed, the finiteness of the space of mechanisms is too strong to be true with general payments that we consider.

Mishra and Quadir [2014] study single-item auctions with a property which Thomson [2016] terms “non-monetary nonbossiness”: the property states that no agent should be able to change other agents’ outcomes without changing their own outcome, but puts no restrictions on payments. Mishra and Quadir [2014] show an allocation rule is implementable in dominant strategies and non-monetary nonbossy if and only if it is strongly rationalizable. Nath and Sen [2015] study “allocation nonbossy” social choice functions, which is the same as non-monetary nonbossiness. They show

that given some richness assumptions, any social choice function that is onto, strategyproof and non-monetary nonbossy must be an affine maximizer. Mukherjee [2015] studies “nonbossiness in decision”, which is again equivalent to non-monetary nonbossiness. The author shows that in the single-item setting, any anonymous, strategyproof, and non-monetary nonbossy allocation rule must be that of a Vickrey auction with a reserve price. As noted by Thomson [2016], non-monetary nonbossiness is incomparable to the notion of nobossiness that we study. Moreover, it is not clear what semantic interpretations (e.g., robustness against a certain form of strategic behavior) non-monetary nonbossiness has beyond its very definition.

Finally, perhaps most closely related to our results is the recent work by Pycia and Raghavan [2021], who study the same notion of nonbossiness as ours in single-item auctions with general payments. They show that (1) the first-price auction with no reserve is the essentially unique mechanism that is nonbossy, individually rational, and efficient in equilibrium, and (2) the first-price auction with the optimal reserve price is the essentially unique mechanism that is nonbossy, individually rational, and revenue maximizing. These results are not directly comparable to ours: aiming to capture robustness against strategic behavior with almost-vanishing externalities, we focus on nonbossy mechanisms that are also incentive-compatible, which first-price auctions clearly are not. In fact, both the results by Pycia and Raghavan [2021] and our results imply that no incentive-compatible and nonbossy mechanism can be efficient (i.e., welfare-maximizing) or revenue-maximizing in the single-item setting.

Stronger notions of robustness. Conceptually, our results are along the line of research on stronger notions of robustness against strategic behavior in mechanism design. Perhaps the closest notion of robustness to ours is obvious strategyproofness by Li [2017] (as well as variants thereof, such as strong obvious strategyproofness [Pycia and Troyan, 2018]), which, roughly speaking, requires that the worst thing that may happen to an agent (over other agents’ actions) under truthful reporting must be at least as good as the best thing that may happen (again, over other agents’ actions) under any possible deviation. This is similar to our notion of nonbossiness at a high level, in the sense that both notions put limitations on how an agent may influence the utility of other agents. However, as we show in Section B, these two notions of robustness are not directly comparable, and in particular, neither notion subsumes the other. Another notion of robustness is credibility by Akbarpour and Li [2020], which roughly says that the principal has no incentive to cheat in the mechanism. In contrast, we assume full commitment power for the principal, and our notion of robustness concerns agents’ incentives only.

Mechanisms with (non-vanishing) externalities. The concept of “externalities” is well established in the economics literature, and there is a significant line of existing work on mechanism design with externalities [Bartling and Netzer, 2016, Bernstein and Winter, 2012, Jehiel et al., 1996, 1999, Segal, 1999], in addition to closely related lines of work on mechanism design with altruistic [Fehr and Schmidt, 2006, Levine, 1998] or spiteful agents [Brandt et al., 2005, Brandt and Weiß, 2001, Morgan et al., 2003, Zhou and Lukose, 2007]. As far as we are aware, we are the first to specifically study the setting of externalities that tend to zero in size (or mechanisms robust to small externalities), and the techniques used for larger externalities are significantly different than the techniques we employ.

Simple mechanisms and posted-price mechanisms. Another related topic is simple mechanisms, and in particular, posted-price mechanisms. The study of simple mechanisms is driven by the fact that in rich environments (e.g., with multiple items and possibly non-additive valuations), optimal mechanisms in terms of revenue or welfare can be complex [Daskalakis et al., 2017, Hart and Reny, 2015, Manelli and Vincent, 2007, Thanassoulis, 2004] and/or hard to compute [Daskalakis et al., 2014,

Mirroknj et al., 2008]. On the other hand, often there are relatively simple mechanisms, such as posted-price mechanisms, that can achieve good (often constant) approximations to the respective benchmarks. In terms of revenue, it is known that when agents are single-parameter or unit-demand, sequential posted-price auctions achieve a constant fraction of the optimal revenue [Cai et al., 2019, Chawla et al., 2007, 2010, Kleinberg and Weinberg, 2012]. Similar results have been established for additive agents [Babaioff et al., 2014, Cai et al., 2019, Hart and Nisan, 2017, Li and Yao, 2013, Yao, 2015] and even subadditive agents [Cai and Zhao, 2017, Rubinstein and Weinberg, 2018] using slightly less simple mechanisms. As for welfare, it is known that anonymous item-pricing mechanisms achieve an approximation factor of 2 for submodular/XOS agents [Dutting et al., 2020, Feldman et al., 2014], and $O(\log \log m)$ for subadditive agents [Dutting et al., 2020, Dütting et al., 2020]. Our results are most closely related to prior work on posted-price mechanisms: restricted to single-parameter settings, our results show that the class of mechanisms that are incentive-compatible, individually rational and nonbossy generalize sequential posted-price mechanisms, in that incentive-compatible, individually rational and nonbossy mechanisms (corresponding to decision lists with exceptions) are slightly less simple, but strictly more powerful.

2 PRELIMINARIES

2.1 Valuation Spaces, Outcome Spaces, and Mechanisms.

We consider multiagent settings with n agents. Each agent i is associated with a valuation space \mathcal{V}_i and a personal outcome space \mathcal{O}_i . Each valuation $v_i \in \mathcal{V}_i$ maps every personal outcome $o_i \in \mathcal{O}_i$ to a real number $v_i(o_i)$. The joint valuation space \mathcal{V} specifies all possible combinations of valuations, and the joint outcome space $\mathcal{O} \subseteq \prod_i \mathcal{O}_i$ specifies all feasible combinations of personal outcomes.

Example 1. Consider a single-parameter auction setting, where there are $k < n$ identical items for sale, and each agent wants at most one of them. Each \mathcal{O}_i can be either $\{0, 1\}$ (corresponding to deterministic allocations) or $[0, 1]$ (corresponding to randomized allocations). Each valuation space \mathcal{V}_i is isomorphic to \mathbb{R}_+ , since each valuation $v_i \in \mathcal{V}_i$ can be described by a nonnegative real number, which is the value $v_i(1)$ of agent i receiving an item. (Abusing notation, we represent each $v_i \in \mathcal{V}_i$ by $v_i(1)$ and simply say $\mathcal{V}_i = \mathbb{R}_+$ in the rest of the paper.) Suppose we consider deterministic mechanisms. Then the joint outcome space is defined to be

$$\mathcal{O} = \{o \in \{0, 1\}^n \mid \|o\|_1 \leq k\}.$$

This captures the fact that no more than k agents can receive an item.

A mechanism $M = (f, p)$ maps each combination of valuations $v = (v_1, \dots, v_n) \in \mathcal{V}$ to a joint outcome $f(v) \in \mathcal{O}$, and charges payments $p(v) \in \mathbb{R}_+^n$. Let $f_i(v)$ denote the personal outcome for agent i in $f(v)$, and $f_{-i}(v)$ all other personal outcomes. Similarly define p_i, p_{-i} . We also generally use $o \in \mathcal{O}$ and $q \in \mathbb{R}_+^n$ to denote specific outcome/payment vectors, as opposed to the mappings $f : \mathcal{V} \rightarrow \mathcal{O}$ and $p : \mathcal{V} \rightarrow \mathbb{R}_+^n$. For each vector of valuations $v \in \mathcal{V}$ and each agent i , we use v_{-i} to denote the valuations of all agent except i . We similarly define $o_{-i}, q_{-i}, \mathcal{V}_{-i}$, etc.

2.2 Incentive-Compatibility and Individual Rationality

The classical notion of (dominant-strategy) individual rationality is captured by the following definition in our model.

Definition 1 (Individual Rationality (IR)). A mechanism $M = (f, p)$ is individually rational (IR) if for all i and $v \in \mathcal{V}$,

$$v_i(f_i(v)) - p_i(v) \geq 0.$$

Now we define incentive-compatiblity. For simplicity, we assume consistent tiebreaking among personal outcomes for each agent i , i.e., there is a strict total order \prec_i^{tb} over O_i such that whenever two outcome-payment pairs induce the same utility, the agent always prefers the one with a “smaller” outcome. We assume this tiebreaking rule is fixed and does not depend on i ’s valuation.² For example, an agent may always prefer receiving an item to not, whenever the two options lead to the same utility (i.e., 0). Then, the classical notion of (dominant-strategy) incentive-compatiblity translates to the following definition.

Definition 2 (Incentive-Compatibility (IC)). A mechanism $M = (f, p)$ is incentive-compatible (IC) if for all $i \in [n]$, $v = (v_i, v_{-i}) \in \mathcal{V}$, and $v'_i \in \mathcal{V}_i$,

$$\begin{aligned} &v_i(f_i(v_i, v_{-i})) - p_i(v_i, v_{-i}) > v_i(f_i(v'_i, v_{-i})) - p_i(v'_i, v_{-i}), \text{ or} \\ &v_i(f_i(v_i, v_{-i})) - p_i(v_i, v_{-i}) = v_i(f_i(v'_i, v_{-i})) - p_i(v'_i, v_{-i}) \text{ and } f_i(v_i, v_{-i}) \leq_i^{\text{tb}} f_i(v'_i, v_{-i}). \end{aligned}$$

2.3 Incentives in the Presence of Almost-Vanishing Externalities

We now formally define agents’ incentives in the presence of almost-vanishing externalities. For each agent i , let \prec_i be i ’s preference over all possible combinations of outcomes and payments, i.e., over $O \times \mathbb{R}_+^n$. This preference is determined by $v_i \in \mathcal{V}_i$ and almost-vanishing externalities together. We assume that \prec_i can be decomposed into an internal component \prec_i^{int} over $O_i \times \mathbb{R}_+$, and an external component (capturing externalities) \prec_i^{ext} over $O_{-i} \times \mathbb{R}_+^{n-1}$, i.e., $\prec_i = (\prec_i^{\text{int}}, \prec_i^{\text{ext}})$. For any two combinations of outcomes and payments (o, q) and $(o', q') \in O \times \mathbb{R}_+^n$, $(o, q) \leq_i (o', q')$ iff: $(o_i, q_i) \prec_i^{\text{int}} (o'_i, q'_i)$, or $(o_i, q_i) = (o'_i, q'_i)$ and $(o_{-i}, q_{-i}) \leq_i^{\text{ext}} (o'_{-i}, q'_{-i})$. That is, the internal component decides the agent’s preference between two combinations, unless the internal outcome-payment pairs are exactly the same, in which case the external component decides the preference. Conceptually, the external component captures externalities, and the fact that it is only used for tiebreaking captures its almost-vanishing nature.

The internal component of the preference $\prec_i^{\text{int}} = \prec_i^{\text{int}}(v_i)$ is induced by i ’s valuation v_i , together with the tiebreaking rule \prec_i^{tb} when two outcome-payment pairs induce the same utility. Formally, $(o_i, q_i) \prec_i^{\text{int}} (o'_i, q'_i)$, iff $v_i(o_i) - q_i > v_i(o'_i) - q'_i$, or $v_i(o_i) - q_i = v_i(o'_i) - q'_i$ and $o_i \prec_i^{\text{tb}} o'_i$.³

We place no restrictions on the external component of the preference \prec_i^{ext} : we allow each \prec_i^{ext} to be any ordering over the outcomes $O \times \mathbb{R}_+^n$ of other agents. Moreover, for every possible ordering \prec over $O \times \mathbb{R}_+^n$, we assume it is possible that $\prec_i^{\text{ext}} = \prec$. That is, we wish to design mechanisms that can handle any possible externality over other agents that an agent possesses.⁴

We now describe the behavior of agents. Each agent i , given i ’s valuation v_i (and therefore the internal component of i ’s preference \prec_i^{int}), all other agents’ valuations v_{-i} , and the external component of i ’s preference \prec_i^{ext} , reports a possibly nontruthful valuation $v'_i \in \mathcal{V}_i$ to the mechanism, where the goal is to achieve a most preferable combination of outcomes and payments according to \prec_i . As in environments without externalities, the exact behavior of agents also depends on what they know about each other. In the rest of this paper, we focus on dominant-strategy robustness against strategic behavior, which means an agent never gets more desirable combinations of outcomes and payments by misreporting their private valuation, no matter what other agents do (in particular, without loss of generality we can assume agents know the valuations and preferences of all other agents). This is captured by the following definition.

²This assumption is unnecessary for some of our results, but without it the proofs become much more tedious.

³Note that given this definition, one can rephrase the definition of incentive-compatiblity in the following way: a mechanism $M = (f, p)$ is incentive-compatible, if for all i , v and v'_i , $(f_i(v_i, v_{-i}), p_i(v_i, v_{-i})) \leq_i^{\text{int}} (f_i(v'_i, v_{-i}), p_i(v'_i, v_{-i}))$.

⁴In Appendix A we show this assumption can in fact be significantly relaxed.

Definition 3 (Robustness with Almost-Vanishing Externalities (RwAVE)). A mechanism $M = (f, p)$ is robust with almost-vanishing externalities, if for any agent i , valuation vector $v \in \mathcal{V}$, public component of i 's preference \prec_i^{ext} , and possible deviation $v'_i \in \mathcal{V}_i$,

$$(f(v_i, v_{-i}), p(v_i, v_{-i})) \leq_i (f(v'_i, v_{-i}), p(v'_i, v_{-i})),$$

where $\leq_i = (\prec_i^{\text{int}}(v_i), \prec_i^{\text{ext}})$.

Note that RwAVE is a stronger notion of robustness than IC defined above, since for any two combinations of outcomes and payments (o, q) and (o', q') ,

$$(o, q) \leq_i (o', q') \implies (o_i, q_i) \leq_i^{\text{int}} (o'_i, q'_i).$$

PROPOSITION 1. *Any RwAVE mechanism M is also IC.*

3 NONBOSSY MECHANISMS

In this section, we introduce the notion of nonbossy mechanisms, and show that it tightly characterizes the class of mechanisms that are robust with almost-vanishing externalities.

Definition 4 (Nonbossiness (NB)). A mechanism $M = (f, p)$ is nonbossy (NB) if for all i , $v = (v_i, v_{-i}) \in \mathcal{V}$ and $v'_i \in \mathcal{V}_i$,

$$\begin{aligned} \text{if: } & f_i(v_i, v_{-i}) = f_i(v'_i, v_{-i}) \text{ and } p_i(v_i, v_{-i}) = p_i(v'_i, v_{-i}) \\ \text{then: } & f(v_i, v_{-i}) = f(v'_i, v_{-i}) \text{ and } p(v_i, v_{-i}) = p(v'_i, v_{-i}). \end{aligned}$$

In words, the above definition says that a mechanism is nonbossy if no agent can change another agent's personal outcome or payment without changing their own personal outcome or payment. Below we show that the semantics of nonbossiness extends well beyond the above definition – in fact, restricted to IC mechanisms, the family of nonbossy mechanisms is precisely the family of mechanisms that are robust with almost-vanishing externalities.

THEOREM 1 (SEMANTICS OF NONBOSSINESS). *A mechanism is IC and NB if and only if it is RwAVE.*

PROOF SKETCH. ⁵ If a mechanism is nonbossy, it is impossible for an agent's deviation to modify the external component of their preference without also modifying the internal component of their preference. Since in IC mechanisms it is impossible for an agent to deviate and improve their internal preference, this means that nonbossy IC mechanisms are RwAVE.

On the other hand, if a mechanism is not nonbossy, then there is a non-truthful deviation for some agent i which modifies the outcomes of the other agents but not i . Such a mechanism cannot be RwAVE, since it is possible that agent i prefers this deviation in the external component of their preference. \square

Deterministic vs. randomized mechanisms. We do not explicitly distinguish between deterministic and randomized mechanisms, as randomization can be captured by extending the outcome space to include all convex combinations of outcomes. Nevertheless, we remark that nonbossiness is generally easy to achieve with randomization: for example, when each $\mathcal{V}_i = \mathbb{R}_+$ and each $\mathcal{O}_i = [0, 1]$ (which correspond to single-parameter environments), a mechanism (f, p) is nonbossy as long as each $f_i(v_i, v_{-i})$ is strictly monotone in v_i . On the other hand, recall that Myerson's characterization [Myerson, 1981] states that any IC mechanism in single-parameter environments must be weakly monotone. Therefore, one can easily adapt any IC mechanism to be also nonbossy with negligible loss in the following way: run the original IC mechanism with probability $1 - \varepsilon$, and any strictly monotone IC mechanism with probability ε , where $\varepsilon > 0$ is arbitrarily small. In light of this, we

⁵For the sake of brevity, we defer most full proofs to Appendix C. Where instructive, we supply proof sketches in their place.

focus on deterministic mechanisms in the rest of the paper: for example, we consider $O_i = \{0, 1\}$ rather than $[0, 1]$ in single parameter environments. [Renato: Is there a sentence or two we can add saying why this is not a satisfactory way to achieve non-bossiness?] [Jon: The only reason I can think of is the generic set of reasons why people dislike randomized mechanisms (tricky to implement, need to trust source of randomness, etc.).]

4 PAYMENT CHARACTERIZATION FOR NONBOSSY MECHANISMS

In this section, we study the basic structure of nonbossy mechanisms. Specifically, whenever an environment has an expressive enough class of valuation functions, we provide a strong characterization of the payment rules of nonbossy mechanisms in this environment. We mathematically formalize this notion of expressivity via what we call the “upper semilattice property”, defined below.

Definition 5 (Upper Semilattice Property). A pair (\mathcal{V}_i, O_i) has the upper semilattice property if for any $o_i \in O_i$, $v_i, v'_i \in \mathcal{V}_i$, there exists $v''_i \in \mathcal{V}_i$ such that for all $o'_i \in O_i$,

$$v''_i(o_i) - v''_i(o'_i) \geq \max\{v_i(o_i) - v_i(o'_i), v'_i(o_i) - v'_i(o'_i)\}.$$

We say v''_i is a common upper bound of v_i and v'_i with respect to o_i .

While the upper semilattice property may appear strong and/or counterintuitive, it turns out that most natural settings commonly studied in the context of mechanism design in fact have this property. For example, it is straightforward to check that in single-parameter settings (where a valuation $v_i \in \mathcal{V}_i$ for a single agent is parameterized by an arbitrary nonnegative real number) Definition 5 holds (e.g., if o_i is allocation of the item and o'_i is non-allocation, we can take $v''_i = \max(v_i, v'_i)$). In fact, the upper semilattice property holds for all the following settings:

- **Common outcome, complete domain:** $O_1 = O_2 = \dots = O_n$, and $O = \{(o_c, \dots, o_c) \mid o_c \in O_1 = \dots = O_n\}$. Each \mathcal{V}_i is the collection of all functions from O_i to \mathbb{R}_+ . This corresponds to settings where the mechanism makes a single public decision that simultaneously affects all agents.
- **Single-parameter agents:** $O_1 = O_2 = \dots = O_n = \{0, 1\}$, and $O \subseteq \times_i O_i$. Each $v_i \in \mathcal{V}_i$ is described by a nonnegative number x_i , where $v_i(1) = x_i$ and $v_i(0) = 0$. This corresponds to settings where the mechanism allocates identical items to agents subject to an arbitrary feasibility constraint, where each agent is interested in at most 1 item.
- **Combinatorial auctions, single-minded agents:** there is a ground set M of items, $O_1 = O_2 = \dots = O_n = 2^M$, and $O \subseteq \times_i O_i$. Each $v_i \in \mathcal{V}_i$ is described by a subset S_i of M and a real number x_i , where $v_i(T) = x_i$ if $S_i \subseteq T$ and $v_i(T) = 0$ otherwise. This corresponds to settings where the mechanism allocates heterogeneous items to agents subject to an arbitrary feasibility constraint, where each agent is only interested in getting all items in a certain set.
- **Combinatorial auctions, valuations “between” additive and XOS:** there is a ground set M of items, $O_1 = O_2 = \dots = O_n = 2^M$, and $O \subseteq \times_i O_i$. Each \mathcal{V}_i is simultaneously a superset of all monotone additive valuations and a subset of all monotone XOS valuations. This corresponds to settings where the mechanism allocates heterogeneous items to agents subject to an arbitrary feasibility constraint, where possible valuations are rich enough to subsume all additive ones, but never go beyond XOS ones.
- **Combinatorial auctions, valuations “beyond” subadditive:** there is a ground set M of items, $O_1 = O_2 = \dots = O_n = 2^M$, and $O \subseteq \times_i O_i$. Each \mathcal{V}_i is simultaneously a superset of all subadditive valuations. This corresponds to settings where the mechanism allocates heterogeneous items to agents subject to an arbitrary feasibility constraint, where possible valuations are rich enough to subsume all monotone subadditive ones.

- **Metric space:** there is a metric space (X_i, d_i) for each agent i where $O_i \subseteq X_i$. Each $v_i \in \mathcal{V}_i$ is induced by a point $x_i \in X_i$, such that $v_i(o_i) = -d_i(x_i, o_i)$. This corresponds to settings where each agent is located in a metric space, which also contains all possible outcomes, and the agent wants the distance to the outcome to be as small as possible (e.g., in facility location games).

PROPOSITION 2. *All the above settings have the upper semilattice property, and therefore admit the payment characterization.*

We now present our payment characterization for nonbossy mechanisms in such environments. Our characterization states that whenever the environment has the upper semilattice property, then for any IC and NB mechanism, the payment rule can be written as a function of the outcome vector only.

THEOREM 2 (PAYMENT CHARACTERIZATION). *If for all $i \in [n]$, (\mathcal{V}_i, O_i) has the upper semilattice property, then for any IC and NB mechanism $M = (f, p)$,*

$$p(v_1, \dots, v_n) = p(v'_1, \dots, v'_n) \quad \text{if} \quad f(v_1, \dots, v_n) = f(v'_1, \dots, v'_n).$$

PROOF. Fix any IC and NB mechanism (f, p) . Consider any two valuation profiles v_1, \dots, v_n and v'_1, \dots, v'_n where

$$f(v_1, \dots, v_n) = f(v'_1, \dots, v'_n) = (o_1, \dots, o_n).$$

For each $i \in [n]$, let v''_i be a common upper bound of v_i and v'_i , satisfying for any $o'_i \in O_i \setminus \{o_i\}$,

$$v''_i(o_i) - v''_i(o'_i) \geq \max\{v_i(o_i) - v_i(o'_i), v'_i(o_i) - v'_i(o'_i)\}.$$

We inductively argue that for any $i \in [n]$,

$$f(v_1, \dots, v_n) = f(v''_1, \dots, v''_i, v_{i+1}, \dots, v_n) \quad \text{and} \quad f(v'_1, \dots, v'_n) = f(v''_1, \dots, v''_i, v'_{i+1}, \dots, v'_n),$$

and

$$p(v_1, \dots, v_n) = p(v''_1, \dots, v''_i, v_{i+1}, \dots, v_n) \quad \text{and} \quad p(v'_1, \dots, v'_n) = p(v''_1, \dots, v''_i, v'_{i+1}, \dots, v'_n).$$

In particular, this implies

$$p(v_1, \dots, v_n) = p(v''_1, \dots, v''_i) = p(v'_1, \dots, v'_n).$$

To show this, we only need to argue that if agent i changes their reported type from v_i to v''_i , then their outcome does not change, i.e.,

$$f_i(v''_1, \dots, v''_{i-1}, v_i, v_{i+1}, \dots, v_n) = f_i(v''_1, \dots, v''_{i-1}, v''_i, v_{i+1}, \dots, v_n).$$

If this is the case, then IC implies that agent i must receive the same payment under both types, i.e.,

$$p_i(v''_1, \dots, v''_{i-1}, v_i, v_{i+1}, \dots, v_n) = p_i(v''_1, \dots, v''_{i-1}, v''_i, v_{i+1}, \dots, v_n),$$

and NB implies that the payments and outcomes do not change for any other agent (which was the inductive claim to be proved). Consider the menu agent i faces when the other agents' valuations are $(v''_1, \dots, v''_{i-1}, v_{i+1}, \dots, v_n)$. When i 's valuation function is v_i , o_i is the utility-maximizing option. Moreover, by the construction of v''_i , when i 's valuation function is v''_i , o_i can only become even more desirable compared to other options, so IC ensures that

$$f_i(v''_1, \dots, v''_i, v_{i+1}, \dots, v_n) = o_i.$$

The other part of the argument (for v') is completely symmetric. □

Necessity of the upper semilattice property. One may naturally wonder if the upper semilattice property is necessary for the characterization – below we show that it in fact is, even if we further restrict our attention to IR mechanisms.

PROPOSITION 3. *There exists an IC, NB, and IR mechanism $M = (f, p)$ where the above payment characterization fails.*

PROOF. Let $n = 2$, $\mathcal{V}_1 = \{x_1, x_2\}$, $\mathcal{V}_2 = \{y_1, y_2\}$, $\mathcal{O}_1 = \mathcal{O}_2 = \{o_1, o_2, o_3\}$, and $\mathcal{O} = \{(o_1, o_1), (o_2, o_2), (o_3, o_3)\}$. Let the valuations be such that

	o_1	o_2	o_3
x_1	1	0	1
x_2	1	1	0
y_1	1	1	0
y_2	1	0	1

Moreover, both agents prefer o_3 to o_1 , and o_1 to o_2 in tiebreaking. Consider the following mechanism.

- $f(x_1, y_1) = (o_1, o_1)$, $p(x_1, y_1) = (1, 0)$,
- $f(x_2, y_2) = (o_1, o_1)$, $p(x_2, y_2) = (0, 1)$,
- $f(x_2, y_1) = (o_2, o_2)$, $p(x_2, y_1) = (0, 0)$,
- $f(x_1, y_2) = (o_3, o_3)$, $p(x_1, y_2) = (0, 0)$.

This does not satisfy the payment characterization, because when the outcome vector is (o_1, o_1) , the payment vector can either be $(1, 0)$ or $(0, 1)$. On the other hand, observe that the mechanism is NB because the (common) outcome changes whenever either agent’s valuation changes, and verifying that the mechanism is IC simply involves checking all relevant cases: when agent 2 reports y_1 , agent 1 would not misreport x_1 as x_2 , because that would create a tie between o_1 and o_2 where agent 1 prefers o_1 . Agent 1 also would not misreport x_2 as x_1 , because that would decrease their utility from 1 to 0. When agent 2 reports y_2 , agent 1 would not misreport x_1 as x_2 , because that would create a tie between o_1 and o_3 where agent 1 prefers o_3 . Agent 1 also would not misreport x_2 as x_1 , because that would decrease their utility from 1 to 0. The case of agent 2 is symmetric. \square

5 FULL CHARACTERIZATIONS IN SPECIFIC SETTINGS

In this section, we present full characterizations of IC, NB, and possibly IR mechanisms in several specific settings.

5.1 Gibbard-Satterthwaite Style Characterization for the Common Outcome Setting

First we present a characterization for the setting where all agents share a common outcome, as described in Section 4. Somewhat surprisingly, this characterization shows that the seemingly rich space of IC and NB mechanisms (which, notably, involve payments) contains only “trivial mechanisms”. These mechanisms must have the same form as the characterization of non-manipulable social choice rules in the celebrated result of Gibbard-Satterthwaite [Gibbard, 1973, Satterthwaite, 1975]: they are either dictatorships, or only two outcomes are relevant.

THEOREM 3. *Consider the setting where all agents share a common outcome and valuations are from the complete domain, i.e., $\mathcal{O} = \{(o_c, \dots, o_c) \mid o_c \in \mathcal{O}_1 = \dots = \mathcal{O}_n\}$ where $|\mathcal{O}_i| < \infty$, and each \mathcal{V}_i is the collection of all functions from \mathcal{O}_i to \mathbb{R}_+ . In this setting, any IC and NB mechanism satisfies the Gibbard-Satterthwaite characterization, i.e., either it is dictatorship or it uses only two outcomes. Formally, any IC and NB mechanism $M = (f, p)$ in this setting must satisfy at least one of the following two conditions:*

- *Dictatorship: there exists an agent i such that $(f(v), p(v)) = (f(v_i), p(v_i))$, i.e., f and p depend only on agent i ’s valuation.*

- *Two outcomes: the image of (f, p) has cardinality at most 2, i.e., $|\{(f(v), p(v)) \mid v \in \mathcal{V}\}| \leq 2$.*

PROOF SKETCH. If the mechanism M is NB, then by our payment characterization, the payment p is solely a function of the eventual outcome o . Consider the social choice problem (where agents wish to elect a single outcome) where the preferences of agent i over the $|\mathcal{O}|$ outcomes are given by the values of $v_i(o) - p_i(o)$ in increasing order. Then M is a mechanism that solves this social choice problem; moreover, we can show that if M is IC for the original problem, it must be IC for this social choice problem, and that every possible collection of n orderings over the outcomes can occur. It follows from the Gibbard-Satterthwaite theorem that M must take the above form. \square

Given the equivalence between RwaVE and nonbossiness (Theorem 1), the above result can be interpreted in the following way: when choosing a common outcome, only trivial mechanisms can be robust against strategic behavior if agents care (even infinitesimally) about each other's utilities.

5.2 Characterizations for Single-Parameter Environments

We now turn to single-parameter environments.

THEOREM 4. *Consider the single-item setting, i.e., $\mathcal{O} = \{o \in \{0, 1\}^n \mid \|o\|_1 = 1\}$ and each $v_i \in \mathcal{V}_i$ is described by a single nonnegative number, namely the value of the item $v_i(1)$. In this setting, any IC, IR, and NB mechanism is a sequential posted-price mechanism, i.e., the mechanism approaches all agents in a predetermined order, makes a (possibly personalized) take-it-or-leave-it offer to each agent, and allocates the item to the first agent who accepts the offer.*

It is easy to see that any sequential posted-price mechanism is IC, IR, and NB. Conceptually, the above characterization says that sequential posted-price mechanisms are the only format of single-item auctions that are robust in the presence of almost-vanishing externalities. Theorem 4 is a corollary of the following characterization of nonbossy mechanisms in general single-parameter environments.

THEOREM 5. *Consider any single-parameter setting, where $\mathcal{O} \subseteq \{0, 1\}^n$ and each $v_i \in \mathcal{V}_i$ is described by a single nonnegative number, namely the value of receiving an item $v_i(1)$. In this setting, any IC, IR, and NB mechanism $M = (f, p)$ is a "decision list" of the following form: each feasible outcome $o \in \mathcal{O}$ is associated with a payment vector $p(o)$, where $p_i(o) = 0$ if $o_i = 0$ for each i , and a set $E(o) \subseteq \mathcal{O}$ of feasible outcomes (we say $E(o)$ is the exception list of o). We say a valuation vector v satisfies an outcome o if for all i , $v_i(o_i) \geq p_i(o)$.⁶ We choose a linear order among all feasible outcomes, and the mechanism checks these outcomes in that order, and outputs the first outcome o satisfied by v such that no feasible outcome $o' \in E(o)$ is also satisfied.*

Formally, for any IC, IR and NB mechanism $M = (f, p)$, there exists a linear order of outcomes $o^{(1)}, \dots, o^{(k)}$ where $k = |\mathcal{O}|$, such that for all $v \in \mathcal{V}$, $f(v) = o^{(j)}$ and $p(v) = p(o^{(j)})$, where

$$j = \min\{j' \in [k] : v \text{ satisfies } o^{(j')} \wedge (\forall o \in E(o^{(j')}), v \text{ does not satisfy } o)\}.$$

Moreover, there exists a way to choose the exception lists such that the order does not matter, i.e., for any valuation vector v , there is precisely one j such that

$$v \text{ satisfies } o^{(j)} \wedge (\forall o \in E(o^{(j)}), v \text{ does not satisfy } o).$$

Unlike in the single-item case, the single-parameter characterization is considerably more general than sequential posted-price mechanisms. In particular, IC, IR, and NB mechanisms allow posting a vector of prices to all agents simultaneously, which is accepted if and only if each agent accepts

⁶Here (and in the rest of the paper) we assume all agents prefer receiving an item when indifferent. If an agent i prefers not receiving an item, one can replace $v_i(o_i) \geq p_i(o)$ with $v_i(o_i) > p_i(o)$ in the above definition and modify the proofs accordingly.

their personal component of the price vector. In fact, as we will show later, mechanisms that are IC, IR, and NB simultaneously are strictly more powerful in terms of efficiency and revenue than sequential posted-price mechanisms. Before diving into the proofs, we first present an example illustrating what these decision lists look like.

Example 2. Consider an environment with $n = 3$ agents and $\mathcal{O} = \{0, 1\}^3$. Consider the following mechanism: allocate to all 3 agents if they all have value at least 1; otherwise, allocate to the “clockwise first” agent with value at least 1 if there exists one; otherwise, do not allocate at all. The full mechanism is described by the table below (the part corresponding to allocating to the “clockwise first” agent is highlighted in red).

(v_1, v_2, v_3)	(o_1, o_2, o_3)	(p_1, p_2, p_3)
$(\geq 1, \geq 1, \geq 1)$	$(1, 1, 1)$	$(1, 1, 1)$
$(\geq 1, \geq 1, < 1)$	$(1, 0, 0)$	$(1, 0, 0)$
$(\geq 1, < 1, \geq 1)$	$(0, 0, 1)$	$(0, 0, 1)$
$(< 1, \geq 1, \geq 1)$	$(0, 1, 0)$	$(0, 1, 0)$
$(\geq 1, < 1, < 1)$	$(1, 0, 0)$	$(1, 0, 0)$
$(< 1, \geq 1, < 1)$	$(0, 1, 0)$	$(0, 1, 0)$
$(< 1, < 1, \geq 1)$	$(0, 0, 1)$	$(0, 0, 1)$
$(< 1, < 1, < 1)$	$(0, 0, 0)$	$(0, 0, 0)$

One may check that the above mechanism is IC, IR and NB. The mechanism corresponds to the following decision list (which is not unique):

- $o^{(1)} = (1, 1, 1)$, $p(o^{(1)}) = (1, 1, 1)$, $E(o^{(1)}) = \emptyset$.
- $o^{(2)} = (1, 0, 0)$, $p(o^{(2)}) = (1, 0, 0)$, $E(o^{(2)}) = \{o^{(3)} = (0, 0, 1)\}$.
- $o^{(3)} = (0, 0, 1)$, $p(o^{(3)}) = (0, 0, 1)$, $E(o^{(3)}) = \{o^{(4)} = (0, 1, 0)\}$.
- $o^{(4)} = (0, 1, 0)$, $p(o^{(4)}) = (0, 1, 0)$, $E(o^{(4)}) = \{o^{(2)} = (1, 0, 0)\}$.
- $o^{(5)} = (0, 0, 0)$, $p(o^{(5)}) = (0, 0, 0)$, $E(o^{(5)}) = \emptyset$.

In particular, exception lists are necessary in this example: there is no way to order $o^{(2)}$, $o^{(3)}$ and $o^{(4)}$ to always choose the clockwise first agent with value at least 1 without exception lists. Also note that the above exception lists are order-specific, i.e., reordering the outcomes may result in a decision list implementing a different mechanism. However, one can make the exception lists order-oblivious by adding $o^{(1)}$ to the exception lists of all other outcomes, and $o^{(2)}$, $o^{(3)}$ and $o^{(4)}$ to the exception list of $o^{(5)}$.

We now describe some of the ideas in the proofs of Theorems 4 and 5. We first sketch how Theorem 5 implies Theorem 4.

PROOF SKETCH OF THEOREM 4. By Theorem 5, we know any nonbossy mechanism for the single-item setting can be written as a decision list mechanism. To show it is a posted-price mechanism, we need to show that this decision list mechanism can be written in such a way that all the exception lists are empty.

To do this, we make use of the fact that nonbossiness imposes some additional strong constraints on the structure of the extension lists. For example, in this single-item setting we can show there cannot exist two outcomes o and o' (where agents i and i' get allocated the items) such that $o \in E(o')$ and $o' \in E(o)$. To see why, assume that o comes before o' in the decision list. Then if outcome o occurs, i' can cause a different outcome to occur by increasing their value and causing o' to be feasible (all the while, o' will never be selected since o is feasible and $o \in E(o)$). By applying similar logic, we can construct a total ordering over all the outcomes (where o dominates o' if

whenever o is feasible, o' cannot be picked) and show that there exists a posted-price mechanism which presents the items in this order. See the full proof for details. \square

We now discuss our proof of Theorem 5. Our proof crucially relies on the following lemma, which shows that if two valuation vectors v and v' both satisfy (in the sense of Theorem 5) outcomes o and o' , then it is not possible that under valuation v the outcome is o and under v' the outcome is o' .

LEMMA 1. *Consider any IC, IR, and NB mechanism $M = (f, p)$ in a single-parameter environment. Fix two different feasible outcomes o and o' , and let $p(o), p(o')$ be the corresponding payment vectors. Choose any two valuation vectors v and v' where both v and v' satisfy both o and o' (i.e., $\min\{v_i, v'_i\} \geq \max\{p_i(o), p_i(o')\}$ for each $i \in [n]$). Then,*

$$f(v) = o \implies f(v') \neq o'.$$

Intuitively, Lemma 1 follows from yet another consequence of nonbossiness and incentive-compatibility; we show that if we increase the valuation from either v or v' to a common maximum v'' , the outcome of the mechanism cannot change without violating one of these two properties.

Lemma 1 is useful for the following reason: instead of having to worry about the behavior of our mechanism $M = (f, p)$ for all valuation vectors v , we can use Lemma 1 to construct a finite number of different “classes” of valuation vector, where M acts identically on every vector in a class.

For example, consider the valuation vector v^O defined via $v_i^O = \max_{o \in O} p(o)_i$; in words, v_i^O is the minimal valuation vector which satisfies every outcome in O . By Lemma 1, if v is a valuation vector which satisfies the outcome $f(v^O)$, then v must also lead to the outcome $f(v^O)$. We can therefore group the set of all valuations v which satisfy $f(v^O)$ into a single “class”.

In the proof of Theorem 4 we repeat this logic for other subsets of outcomes to first construct a structure we call a “decision tree”. In a decision tree, every node is labeled by a subset $S \subseteq O$ of outcomes and corresponds to a class of valuations “equivalent to” (in the sense that we can apply Lemma 1 to argue they have the same outcome) the minimal valuation v^S that satisfies every outcome in S . The root of this tree is labeled by O and contains the class of valuations described above. The children of a node labeled S contain the maximal subsets S' of S where $v^{S'}$ does not belong to the class of valuations of v^S .

By constructing this tree in this way, we can prove that each valuation v is associated to some node, and in fact it is associated to the node with the maximal label S such that v does not satisfy any outcome outside of S . This gives us a “decision-list-esque” mechanism, where for each v we iterate through the nodes of our tree until we find a node S where v both satisfies $f(v^S)$ and does not satisfy any outcome $o \notin S$. This process may consider the same outcomes more than once, but with a little bit more care we can convert this to a single decision list, and even an order-oblivious one (see the full proof for details).

6 SEPARATING POSTED-PRICE MECHANISMS AND NONBOSSY MECHANISMS

We have seen that in general single-parameter environments, the set of nonbossy mechanisms is richer than the set of posted-price mechanisms. However, it is not immediately clear how much additional power nonbossy mechanisms have in terms of optimizing metrics of interest to mechanism designers (e.g. revenue or welfare). In this section, we establish strict separations in terms of revenue and efficiency between mechanisms that are IC, IR and NB, and sequential posted-price mechanisms in single-parameter environments. In particular, we identify two sources of separation: rich feasibility constraints and correlation between agents.

6.1 General Downward-Closed Feasibility Constraints

As Theorem 4 shows, IC, IR and NB mechanisms are equivalent to sequential posted-price mechanisms when there is a single item for sale — that is, when the family of all feasible sets of agents to receive an item contains only singletons. More generally, it is known (see, e.g., [Chawla et al., 2010, Kleinberg and Weinberg, 2012]) that sequential posted-price mechanisms can get at least $1/2$ of the optimal welfare with matroid feasibility constraints, or even intersections of a constant number of matroids⁷. However, when we allow arbitrary downward-closed feasibility constraints, sequential posted-price mechanisms can no longer achieve constant approximations to the optimal welfare. Before proceeding, we first rephrase the notion of downward-closed feasibility constraints in our terminology.

Definition 6 (Downward-Closed Feasibility Constraints). A single-parameter environment has downward-closed feasibility constraints, if the space of feasible outcomes $\mathcal{O} \subseteq \{0, 1\}^n$ satisfies: for any $o \in \mathcal{O}$ and o' where $o'_i \leq o_i$ for each i , it is always the case that $o' \in \mathcal{O}$.

Babaioff et al. [2007] prove the following lower bound for sequential posted-price mechanisms in environments with downward-closed feasibility constraints.

THEOREM 6 ([BABAI OFF ET AL., 2007], REPHRASED). *The best possible approximation ratio that any sequential posted-price mechanism can achieve in single-parameter Bayesian settings with independent agents and downward-closed feasibility constraints, parametrized by the maximum number r of agents who can simultaneously receive an item, a.k.a. the rank of the feasibility constraint, is $\Omega(r)$.*

We will see momentarily that IC, IR and NB mechanisms achieve strictly (and in fact, exponentially) better approximation ratios in the same environments. To be self-contained, we briefly review the hard instance used in [Babaioff et al., 2007], which also provides a general sense of what downward-closed feasibility constraints look like.

Example 3. Suppose the n agents are partitioned into groups of size r . The value of each agent is independently 1 with probability p , and 0 otherwise. The mechanism can allocate to any set of agents, subject to the constraint that all agents receiving an item must be in the same group. Consider the case where r is much smaller than n , and $p = 1/r$. By choosing r appropriately, one can create a situation where with high probability, there exists a group in which all agents have value 1, so the expected optimal welfare is very close to r . On the other hand, for each group, the expected total value is 1, and moreover, this value is smaller than 2 even conditioned on some agent in the group having value 1. In fact, that upper bounds the expected welfare of any sequential posted-price mechanism. This is because when a sequential posted-price mechanism decides to allocate an item to an agent, it is essentially committing to allocating to all agents in this group, and no one else. At this point, the only information the mechanism has about this group is the value of the current agent that the mechanism is visiting, which is at most 1. So in any case, the conditional expected welfare is upper bounded by 2. This creates a gap of $\Omega(r)$.

Below we show that IC, IR and NB mechanisms can achieve an exponentially improved approximation ratio in such environments.

THEOREM 7. *For any single-parameter Bayesian setting with independent agents and downward-closed feasibility constraints, there exists an IC, IR and NB mechanism that achieves an $O(\log r)$ -approximation with respect to the optimal welfare, where r is the rank of the feasibility constraint.*

⁷This is also true for revenue since one can replace the actual value of each agent with the virtual value. Given this reduction, we focus on welfare in the rest of the subsection.

The proof works by reducing to the case where the support of every agent i 's value distribution is $\{0, a_i\}$ for some fixed $a_i \geq 0$. We first show that in that case, there always exists an IC, IR and NB mechanism which always extracts the full optimal welfare as revenue.

LEMMA 2. *For any single-parameter Bayesian setting with independent agents with binary supports and downward-closed feasibility constraints, there exists an IC, IR and NB mechanism that always extracts the full optimal welfare as revenue. Moreover, the revenue of the mechanism is weakly monotone in the values, which is true even if values are not restricted to binary supports.*

PROOF. Consider the following mechanism:

- Let $k = |\mathcal{O}|$. Sort all outcomes in \mathcal{O} into $o^{(1)}, \dots, o^{(k)}$ in decreasing order by the total welfare of the outcome, i.e., such that for each $j \in [k - 1]$,

$$\sum_{i \in [n]} a_i \cdot o_i^{(j)} \geq \sum_{i \in [n]} a_i \cdot o_i^{(j+1)}.$$

- For each $j \in [k]$, post the price vector $p(o^{(j)})$ where for each i ,

$$p_i(o^{(j)}) = a_i \cdot o_i^{(j)}.$$

If the valuation vector v satisfies this outcome $o^{(j)}$, i.e., for each i ,

$$v_i \geq p_i(o^{(j)}),$$

then output outcome $o^{(j)}$ and terminate. Otherwise, proceed to the next outcome.

One can check that the mechanism is in fact IC, IR and NB. To see why the mechanism gets the full optimal welfare, fix a valuation vector v and consider the welfare-maximizing outcome $o^{(j)}$. Without loss of generality, $o^{(j)}$ satisfies: for each i ,

$$o_i^{(j)} = 1 \implies v_i = a_i.$$

This is because the feasibility constraint is downward-closed, so we can set $o_i^{(j)} = 0$ for any agent i with value $v_i = 0$, and the resulting outcome is still feasible. Now v must satisfy $o^{(j)}$, so the above mechanism must output some outcome before (including) $o^{(j)}$. If the mechanism outputs an outcome before $o^{(j)}$, then the welfare can never be smaller, because of the way we sort the outcomes. So, the mechanism must always extract the optimal welfare as revenue with binary value distributions. Finally, it is easy to check the revenue of the mechanism is monotone in the values, since increasing the values can never make a satisfied outcome unsatisfied. \square

We remark that the value distributions in Example 3 in fact have binary supports, and therefore one can apply Lemma 2 to get an IC, IR and NB mechanism that extracts the full optimal welfare as revenue, which is even stronger than the logarithmic guarantee in Theorem 7 for the general case. We are now ready to prove Theorem 7.

PROOF OF THEOREM 7. Without loss of generality, suppose for each agent i , it is feasible to allocate an item to i and no one else (when this is infeasible for some i , then i can never receive an item, so we can ignore i). Let $o^*(v)$ be the welfare-maximizing outcome for a valuation vector v , i.e.,

$$o^*(v) = \operatorname{argmax}_{o \in \mathcal{O}} \sum_i v_i(o_i).$$

Let OPT be the expected optimal welfare, i.e.,

$$\text{OPT} = \mathbb{E}_v \left[\sum_i v_i(o_i^*(v)) \right].$$

We decompose OPT into 3 parts:

$$\begin{aligned}\text{OPT}_1 &= \mathbb{E}_v \left[\sum_i v_i(o_i^*(v)) \cdot \mathbb{I}[v_i(o_i^*(v)) \geq 2\text{OPT}] \right], \\ \text{OPT}_2 &= \mathbb{E}_v \left[\sum_i v_i(o_i^*(v)) \cdot \mathbb{I}[\text{OPT}/(2r) \leq v_i(o_i^*(v)) < 2\text{OPT}] \right], \\ \text{OPT}_3 &= \mathbb{E}_v \left[\sum_i v_i(o_i^*(v)) \cdot \mathbb{I}[v_i(o_i^*(v)) < \text{OPT}/(2r)] \right].\end{aligned}$$

Clearly $\text{OPT} = \text{OPT}_1 + \text{OPT}_2 + \text{OPT}_3$. Also observe that

$$\begin{aligned}\text{OPT}_3 &= \mathbb{E}_v \left[\sum_i v_i(o_i^*(v)) \cdot \mathbb{I}[v_i(o_i^*(v)) < \text{OPT}/(2r)] \right] \\ &= \mathbb{E}_v \left[\sum_{i:o_i^*(v)=1} v_i(o_i^*(v)) \cdot \mathbb{I}[v_i(o_i^*(v)) < \text{OPT}/(2r)] \right] \quad (v_i(0) = 0 \text{ for each } v \text{ and } i) \\ &< \mathbb{E}_v \left[\sum_{i:o_i^*(v)=1} \text{OPT}/(2r) \right] \\ &\leq \mathbb{E}_v [r \cdot \text{OPT}/(2r)] \quad (\text{maximum size of feasible allocation set is } r) \\ &= \text{OPT}/2.\end{aligned}$$

As a result,

$$\text{OPT}_1 + \text{OPT}_2 \geq \text{OPT}/2,$$

and

$$\max\{\text{OPT}_1, \text{OPT}_2\} \geq \text{OPT}/4.$$

Our plan is to get a significant fraction of the larger one between OPT_1 and OPT_2 .

First consider the case where $\text{OPT}_1 \geq \text{OPT}/4$. Consider the following mechanism: visit all agents in an arbitrary order; for each agent i , if i 's value $v_i(1)$ is at least 2OPT , then we allocate to agent i and terminate; otherwise, proceed to the next agent. We argue that this mechanism guarantees welfare at least $\text{OPT}_1/2 \geq \text{OPT}/8$. In fact, for each i , let

$$\alpha_i = \Pr[v_i(1) \geq 2\text{OPT}].$$

Observe that α_i is the probability that we flip a coin for each agent i , and in that case, i 's expected value is $\mathbb{E}[v_i(1) \mid v_i(1) \geq 2\text{OPT}]$. We have

$$\begin{aligned}\sum_i \alpha_i \cdot \mathbb{E}[v_i(1) \mid v_i(1) \geq 2\text{OPT}] &= \mathbb{E} \left[\sum_i v_i(1) \cdot \mathbb{I}[v_i(1) \geq 2\text{OPT}] \right] \\ &\geq \mathbb{E} \left[\sum_i v_i(o_i^*(v)) \cdot \mathbb{I}[v_i(o_i^*(v)) \geq 2\text{OPT}] \right] = \text{OPT}_1.\end{aligned}$$

And moreover, since whenever there is some i such that $v_i(1) \geq 2\text{OPT}$, the optimal welfare is at least 2OPT , the probability of this event cannot be too large. That is,

$$(1 - \prod_i (1 - \alpha_i)) \cdot (2\text{OPT}) \leq \text{OPT} \implies \prod_i (1 - \alpha_i) \geq 1/2.$$

Now come back to our mechanism. Since agents are independent, the expected contribution to the welfare from agent i is

$$\begin{aligned} \prod_{i' < i} (1 - \alpha_{i'}) \cdot \alpha_i \cdot \mathbb{E}[v_i(1) \mid v_i(1) \geq 2\text{OPT}] &\geq \prod_{i' \in [n]} (1 - \alpha_{i'}) \cdot \alpha_i \cdot \mathbb{E}[v_i(1) \mid v_i(1) \geq 2\text{OPT}] \\ &\geq \frac{1}{2} \cdot \alpha_i \cdot \mathbb{E}[v_i(1) \mid v_i(1) \geq 2\text{OPT}]. \end{aligned}$$

Summing over i , the expected welfare we get is at least

$$\frac{1}{2} \sum_i \alpha_i \cdot \mathbb{E}[v_i(1) \mid v_i(1) \geq 2\text{OPT}] \geq \frac{1}{2} \cdot \text{OPT}_1 \geq \text{OPT}/8.$$

Now consider the case where $\text{OPT}_2 \geq \text{OPT}/4$. We further subdivide OPT_2 into about $\log r$ bins, and give a mechanism which gets a constant fraction of the contribution of the largest bin (which contributes $\text{OPT}_2/O(\log r)$ to the welfare) in expectation.⁸ For each valuation vector v and $j \in \{1, \dots, \lceil \log(4r) \rceil\}$, let

$$\text{bin}_j(v) = \{i \in [n] : v_i(o_i^*(v)) \in [2\text{OPT}/2^j, 2\text{OPT}/2^{j-1}]\}.$$

Note that for each v , the union of these bins does not necessarily contain all agents. However, since these bins completely cover the possible range of values that may contribute to OPT_2 , we have

$$\text{OPT}_2 \leq \sum_j \mathbb{E} \left[\sum_{i \in \text{bin}_j(v)} v_i(o_i^*(v)) \right].$$

So there must exist some j^* , such that

$$\mathbb{E} \left[\sum_{i \in \text{bin}_{j^*}(v)} v_i(o_i^*(v)) \right] \geq \text{OPT}_2/O(\log r) = \text{OPT}/O(\log r).$$

We perform the following (imaginary) transformation to the value distributions: for each agent i , consider the distribution which has probability mass $\Pr[v_i \in [2\text{OPT}/2^{j^*}, 2\text{OPT}/2^{j^*-1}]]$ at $2\text{OPT}/2^{j^*}$, and the rest of the probability mass at 0. These imaginary value distributions are binary, and dominated by the actual value distributions. We then run the mechanism constructed in Lemma 2 with these imaginary distributions as input. Since the revenue of the mechanism is weakly monotone in the values, running it with the actual value distributions can only increase its revenue. Moreover, for each v and $i \in \text{bin}_{j^*}(v)$, $v_i \leq 2\text{OPT}/2^{j^*-1}$. So the expected revenue that our mechanism extracts is at least

$$\mathbb{E} \left[\sum_{i \in \text{bin}_{j^*}(v)} 2\text{OPT}/2^{j^*} \cdot o_i^*(v) \right] \geq \mathbb{E} \left[\frac{1}{2} \sum_{i \in \text{bin}_{j^*}(v)} v_i(o_i^*(v)) \right] \geq \text{OPT}/O(\log r).$$

So the welfare guaranteed by the mechanism is $\text{OPT}/O(\log r)$. This concludes the proof. \square

6.2 Correlated Valuations

We now consider another case where IC, IR and NB mechanisms perform strictly better than sequential posted-price mechanisms: when agents may have correlated valuations. In particular, we give an example with only 2 agents, where any sequential posted-price mechanism can only extract $4/5$ of the optimal welfare as revenue, while there is an IC, IR and NB mechanism that extracts the full optimal welfare.

⁸In fact, the mechanism extracts a constant fraction of the welfare from that bin as revenue.

Example 4. Consider the following environment with $n = 2$ agents, and no feasibility constraints (i.e., the mechanism may allocate to any set of agents). The values of the two agents are $(1, 1)$ with probability $1/5$, $(2, 0)$ with probability $2/5$, and $(0, 2)$ with probability $2/5$. Consider any sequential posted-price mechanism, which without loss of generality visits agent 1 first, and then agent 2. The only reasonable ways to price the agents are the following:

- Set the price for agent 1 to 1. If agent 1 buys, then set the price for agent 2 to 1. Otherwise, set the price for agent 2 to 2. The expected revenue of this mechanism is $8/5$. In particular, when the values are $(2, 0)$, the mechanism only extracts revenue 1 from agent 1 (it extracts the full optimal welfare of 2 in all other cases).
- Set the price for agent 1 to 2, and set the price for agent 2 to 1. The expected revenue of this mechanism is $7/5$. In particular, when the values are $(0, 2)$, the mechanism only extracts revenue 1 from agent 2, and when the values are $(1, 1)$, the mechanism only extracts 0 from agent 1 (it extracts the full optimal welfare of 2 when the values are $(2, 0)$).
- Set the price for agent 1 to 2, and set the price for agent 2 to 2. The expected revenue of this mechanism is $8/5$. In particular, when the values are $(1, 1)$, the mechanism does not extract any revenue (it extracts the full optimal welfare of 2 in all other cases).

So the optimal sequential posted-price mechanism achieves revenue $8/5$. On the other hand, consider the following decision list (which is IC, IR and NB):

- $o^{(1)} = (1, 1)$, $p(o^{(1)}) = (1, 1)$, $E(o^{(1)}) = \emptyset$.
- $o^{(2)} = (1, 0)$, $p(o^{(2)}) = (2, 0)$, $E(o^{(2)}) = \emptyset$.
- $o^{(3)} = (0, 1)$, $p(o^{(3)}) = (0, 2)$, $E(o^{(3)}) = \emptyset$.

One can check the decision list extracts revenue 2 in all cases.

REFERENCES

- Mohammad Akbarpour and Shengwu Li. 2020. Credible auctions: A trilemma. *Econometrica* 88, 2 (2020), 425–467.
- Moshe Babaioff, Nicole Immorlica, and Robert Kleinberg. 2007. Matroids, secretary problems, and online mechanisms. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. 434–443.
- Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S Matthew Weinberg. 2014. A simple and approximately optimal mechanism for an additive buyer. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 21–30.
- Sophie Bade. 2020. Random serial dictatorship: the one and only. *Mathematics of Operations Research* 45, 1 (2020), 353–368.
- Björn Bartling and Nick Netzer. 2016. An externality-robust auction: Theory and experimental evidence. *Games and Economic Behavior* 97 (2016), 186–204.
- Shai Bernstein and Eyal Winter. 2012. Contracting with heterogeneous externalities. *American Economic Journal: Microeconomics* 4, 2 (2012), 50–76.
- Felix Brandt, Tuomas Sandholm, and Yoav Shoham. 2005. Spiteful Bidding in Sealed-Bid Auctions. In *Computing and Markets*.
- Felix Brandt and Gerhard Weiß. 2001. Antisocial agents and Vickrey auctions. In *International Workshop on Agent Theories, Architectures, and Languages*. Springer, 335–347.
- Yang Cai, Nikhil R Devanur, and S Matthew Weinberg. 2019. A duality-based unified approach to bayesian mechanism design. *SIAM J. Comput.* 0 (2019), STOC16–160.
- Yang Cai and Mingfei Zhao. 2017. Simple mechanisms for subadditive buyers via duality. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. 170–183.
- Shuchi Chawla, Jason D Hartline, and Robert Kleinberg. 2007. Algorithmic pricing via virtual valuations. In *Proceedings of the 8th ACM Conference on Electronic Commerce*. 243–251.
- Shuchi Chawla, Jason D Hartline, David L Malec, and Balasubramanian Sivan. 2010. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the forty-second ACM symposium on Theory of computing*. 311–320.
- Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. 2014. The complexity of optimal mechanism design. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 1302–1318.
- Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. 2017. Strong Duality for a Multiple-Good Monopolist. *Econometrica* 85, 3 (2017), 735–767.

- Paul Dütting, Michal Feldman, Thomas Kesselheim, and Brendan Lucier. 2020. Prophet inequalities made easy: Stochastic optimization by pricing nonstochastic inputs. *SIAM J. Comput.* 49, 3 (2020), 540–582.
- Paul Dütting, Thomas Kesselheim, and Brendan Lucier. 2020. An $O(\log \log m)$ prophet inequality for subadditive combinatorial auctions. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 306–317.
- Ernst Fehr and Klaus M Schmidt. 2006. The economics of fairness, reciprocity and altruism—experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity* 1 (2006), 615–691.
- Michal Feldman, Nick Gravin, and Brendan Lucier. 2014. Combinatorial auctions via posted prices. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 123–135.
- Allan Gibbard. 1973. Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society* (1973), 587–601.
- Sergiu Hart and Noam Nisan. 2017. Approximate revenue maximization with multiple items. *Journal of Economic Theory* 172 (2017), 313–347.
- Sergiu Hart and Philip J Reny. 2015. Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theoretical Economics* 10, 3 (2015), 893–922.
- John William Hatfield. 2009. Strategy-proof, efficient, and nonbossy quota allocations. *Social Choice and Welfare* 33, 3 (2009), 505–515.
- Philippe Jehiel, Benny Moldovanu, and Ennio Stacchetti. 1996. How (not) to sell nuclear weapons. *The American Economic Review* (1996), 814–829.
- Philippe Jehiel, Benny Moldovanu, and Ennio Stacchetti. 1999. Multidimensional mechanism design for auctions with externalities. *Journal of economic theory* 85, 2 (1999), 258–293.
- Robert Kleinberg and Seth Matthew Weinberg. 2012. Matroid prophet inequalities. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 123–136.
- Fuhito Kojima. 2010. Impossibility of stable and nonbossy matching mechanisms. *Economics Letters* 107, 1 (2010), 69–70.
- David K Levine. 1998. Modeling altruism and spitefulness in experiments. *Review of economic dynamics* 1, 3 (1998), 593–622.
- Shengwu Li. 2017. Obviously strategy-proof mechanisms. *American Economic Review* 107, 11 (2017), 3257–87.
- Xinye Li and Andrew Chi-Chih Yao. 2013. On revenue maximization for selling multiple independently distributed items. *Proceedings of the National Academy of Sciences* 110, 28 (2013), 11232–11237.
- Alejandro M Manelli and Daniel R Vincent. 2007. Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic theory* 137, 1 (2007), 153–185.
- Vahab Mirrokni, Michael Schapira, and Jan Vondrák. 2008. Tight information-theoretic lower bounds for welfare maximization in combinatorial auctions. In *Proceedings of the 9th ACM conference on Electronic commerce*. 70–77.
- Debasis Mishra and Abdul Quadir. 2014. Non-bossy single object auctions. *Economic Theory Bulletin* 2, 1 (2014), 93–110.
- John Morgan, Ken Steiglitz, and George Reis. 2003. The spite motive and equilibrium behavior in auctions. *Contributions in Economic Analysis & Policy* 2, 1 (2003), 1–25.
- Conan Mukherjee. 2015. On Axioms Underlying Use of Reserve Price. *Working Paper/Department of Economics, School of Economics and Management, Lund University* 7 (2015).
- Roger B Myerson. 1981. Optimal auction design. *Mathematics of operations research* 6, 1 (1981), 58–73.
- Swaprava Nath and Arunava Sen. 2015. Affine maximizers in domains with selfish valuations. *ACM Transactions on Economics and Computation (TEAC)* 3, 4 (2015), 1–19.
- Szilvia Pápai. 2001. Strategyproof and nonbossy multiple assignments. *Journal of Public Economic Theory* 3, 3 (2001), 257–271.
- Marek Pycia and Madhav Raghavan. 2021. Non-bossiness and First-Price Auctions. *Available at SSRN 3941784* (2021).
- Marek Pycia and Peter Troyan. 2018. Obvious dominance and random priority. *SSRN Electronic Journal* 2853563 (2018).
- Aviad Rubinfeld and S Matthew Weinberg. 2018. Simple mechanisms for a subadditive buyer and applications to revenue monotonicity. *ACM Transactions on Economics and Computation (TEAC)* 6, 3-4 (2018), 1–25.
- Mark Allen 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 (1975), 187–217.
- Mark A Satterthwaite and Hugo Sonnenschein. 1981. Strategy-proof allocation mechanisms at differentiable points. *The Review of Economic Studies* 48, 4 (1981), 587–597.
- Ilya Segal. 1999. Contracting with externalities. *The Quarterly Journal of Economics* 114, 2 (1999), 337–388.
- Lars-Gunnar Svensson. 1999. Strategy-proof allocation of indivisible goods. *Social Choice and Welfare* 16, 4 (1999), 557–567.
- Lars-Gunnar Svensson and Bo Larsson. 2002. Strategy-proof and nonbossy allocation of indivisible goods and money. *Economic Theory* 20, 3 (2002), 483–502.
- John Thanassoulis. 2004. Hagglng over substitutes. *Journal of Economic theory* 117, 2 (2004), 217–245.
- William Thomson. 2016. Non-bossiness. *Social Choice and Welfare* 47, 3 (2016), 665–696.
- Andrew Chi-Chih Yao. 2015. An n-to-1 bidder reduction for multi-item auctions and its applications. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. 92–109.

Yunhong Zhou and Rajan Lukose. 2007. Vindictive bidding in keyword auctions. In *Proceedings of the ninth international conference on Electronic commerce*. 141–146.

A RESTRICTING POSSIBLE EXTERNALITIES

In our definition of agents with externalities in Section 2.3, we allowed the external component of their preference (the “externality”) to be any arbitrary ordering over the outcomes and payments of the other agents. In this appendix, we show that this freedom is unnecessary – all our results continue to hold with only mild constraints on the set of possible externalities (namely, as long as each agent can express either altruistic or malicious preferences towards other agents, our characterization continues to hold).

We formalize this as follows. Fixing other agents’ valuations to be v_{-i} , let $\mathcal{P}_i^{\text{ext}}(v_{-i})$ be the set of allowed preferences for agent i over other agents’ outcome-payment pairs. For any agent $j \neq i$, we say $\mathcal{P}_i^{\text{ext}}(v_{-i})$ covers $\prec_j^{\text{int}}(v_j)$, if at least one of the following two conditions hold:

- There exists a preference $\prec_i^{\text{ext}} \in \mathcal{P}_i^{\text{ext}}(v_{-i})$, such that for all $(o, q) \in \mathcal{O} \times \mathbb{R}_+^n$, if $(o_j, q_j) \prec_j^{\text{int}}(o'_j, q'_j)$, then $(o_{-i}, q_{-i}) \prec_i^{\text{ext}}(o'_{-i}, q'_{-i})$. When i has such an external component of preference, we say i cares positively about j .
- There exists a preference $\prec_i^{\text{ext}} \in \mathcal{P}_i^{\text{ext}}(v_{-i})$, such that for all $(o, q) \in \mathcal{O} \times \mathbb{R}_+^n$, if $(o_j, q_j) \prec_j^{\text{int}}(o'_j, q'_j)$, then $(o'_{-i}, q'_{-i}) \prec_i^{\text{ext}}(o_{-i}, q_{-i})$. When i has such an external component of preference, we say i cares negatively about j .

We make the following richness assumption on the possible externalities for each agent.

ASSUMPTION 1 (RICHNESS OF EXTERNALITIES). For any agent i and $v_{-i} \in \mathcal{V}_{-i}$, $\mathcal{P}_i^{\text{ext}}(v_{-i})$ covers all $\{\prec_j^{\text{int}}(v_j)\}_{j \neq i}$.

We also remark that Assumption 1 is in a sense necessary for the concept of RwAVE to distinguish itself from IC as a stronger notion of robustness – without the richness assumption, it could be the case that all agents are always indifferent to others’ interests, and as a result RwAVE would be precisely equivalent to IC. With Assumption 1, all of our proofs (most notably our characterization of RwAVE mechanisms as non-bossy in Theorem 1 continue to hold).

B NONBOSSINESS AND OBVIOUS STRATEGYPROOFNESS

The notion of nonbossiness is reminiscent of another important notion of robustness against strategic behavior, namely obvious strategyproofness (OSP), proposed by Li [2017]. Roughly speaking, a mechanism (as an extensive-form game) is OSP, if for any agent, the worst (over other agents’ actions) thing that may happen under truthful reporting is at least as good as the best thing that may happen under any deviation from reporting truthfully. A mechanism, as a pair of allocation and payment rules, is OSP implementable, if there is a way to implement this mechanism using an extensive-form game that is OSP. OSP implementability appears conceptually related to NB, since both notions require that it is “hard” for an agent to manipulate the utility of another agent.

We remark that similar as nonbossiness and OSP appear, neither of the two notions is stronger or weaker than the other, even when restricted to IC and IR mechanisms.

THEOREM 8. *There exists an NB, IC and IR mechanism that is not OSP implementable, and an OSP implementable, IC and IR mechanism that is not NB.*

PROOF. For the former, consider the mechanism used in the proof of Proposition 3, and possible ways to implement it using an extensive-form game. Since there are only 2 agents, each with two possible types, the mechanism should interact with each agent exactly once. Without loss of generality, suppose the mechanism interacts with agent 1 first (the 2 agents are symmetric).

Consider the case where agent 1's valuation is $x_1 \in \mathcal{V}_1$. If agent 1 reports x_1 truthfully, then the worst thing that can happen is agent 2 reporting y_1 , in which case agent 1 gets outcome o_1 and pays 1, leading to a utility of 0. However, if agent 1 deviates and reports x_2 , then the best thing that can happen is agent 2 reporting y_2 , in which case agent 1 gets outcome o_1 and pays 0, leading to a utility of 1. This means the mechanism is not OSP implementable.

For the latter, it is known that the second-price auction is OSP implementable (using the English auction), but it is not NB, since the payment of the winner depends on the highest other bid. \square

Despite this non-comparability, many of the NB mechanisms we study and construct in this paper are indeed also OSP implementable. Specifically, sequential posted-price mechanisms for a single item are OSP in their default implementation, as are incentive-compatible decision list mechanisms with empty exception lists (in particular, the mechanism of Lemma 2 which exponentially outperforms the best sequential posted-price mechanism is of this form).

C OMITTED PROOFS

C.1 Proof of Theorem 1

PROOF OF THEOREM 1. First, fix an environment $(\mathcal{V}, \mathcal{O})$ and a mechanism $M = (f, p)$. Assuming M is IC and NB, we show it is RwAVE. To this end, consider any agent i , valuation vector $v \in \mathcal{V}$, external component of i 's preference $\prec_i^{\text{ext}} \in \mathcal{P}_i^{\text{ext}}(v_{-i})$, and deviation $v'_i \in \mathcal{V}_i$. Also let \prec_i^{int} be the internal component of i 's preference induced by v_i . Since M is IC, we have

$$(f_i(v_i, v_{-i}), p_i(v_i, v_{-i})) \leq_i^{\text{int}} (f_i(v'_i, v_{-i}), p_i(v'_i, v_{-i})).$$

So, in order to show

$$(f(v_i, v_{-i}), p(v_i, v_{-i})) \leq_i (f(v'_i, v_{-i}), p(v'_i, v_{-i})),$$

we only need to argue that whenever

$$(f_i(v_i, v_{-i}), p_i(v_i, v_{-i})) = (f_i(v'_i, v_{-i}), p_i(v'_i, v_{-i})),$$

we have

$$(f_{-i}(v_i, v_{-i}), p_{-i}(v_i, v_{-i})) \leq_i^{\text{ext}} (f_{-i}(v'_i, v_{-i}), p_{-i}(v'_i, v_{-i})).$$

This follows directly from the fact that M is NB, since when

$$(f_i(v_i, v_{-i}), p_i(v_i, v_{-i})) = (f_i(v'_i, v_{-i}), p_i(v'_i, v_{-i})),$$

NB requires that

$$(f(v_i, v_{-i}), p(v_i, v_{-i})) = (f(v'_i, v_{-i}), p(v'_i, v_{-i})),$$

and therefore

$$(f_{-i}(v_i, v_{-i}), p_{-i}(v_i, v_{-i})) \leq_i^{\text{ext}} (f_{-i}(v'_i, v_{-i}), p_{-i}(v'_i, v_{-i})).$$

Now consider the other direction. Assuming M is RwAVE, we show it is IC and NB. First observe that any RwAVE mechanism M is IC. Below we show M is also NB. Fix any agent i , valuation vector v , and deviation v'_i . Suppose towards a contradiction that there exists $j \neq i$, such that

$$(f_j(v_i, v_{-i}), p_j(v_i, v_{-i})) \neq (f_j(v'_i, v_{-i}), p_j(v'_i, v_{-i})).$$

Moreover, without loss of generality,

$$(f_j(v'_i, v_{-i}), p_j(v'_i, v_{-i})) \prec_j^{\text{int}} (f_j(v_i, v_{-i}), p_j(v_i, v_{-i})).$$

Now, let us choose an external preference component \prec_i^{ext} for agent i such that for all $(o, q) \in \mathcal{O} \times \mathbb{R}_+^q$, if $(o_j, q_j) \prec_j^{\text{int}} (o'_j, q'_j)$, then $(o_{-i}, q_{-i}) \prec_i^{\text{ext}} (o'_{-i}, q'_{-i})$. Note that even if we are in the restricted model

of externalities of Appendix A, Assumption 1 guarantees such an ordering exists in $\mathcal{P}_i^{\text{ext}}(v_{-i})$. When i has this external preference component, we have

$$(f(v'_i, v_{-i}), p(v'_i, v_{-i})) <_i (f(v_i, v_{-i}), p(v_i, v_{-i})).$$

However, RWAve requires that

$$(f(v_i, v_{-i}), p(v_i, v_{-i})) \leq_i (f(v'_i, v_{-i}), p(v'_i, v_{-i})),$$

a contradiction. This concludes the proof. \square

C.2 Proof of Proposition 2

PROOF OF PROPOSITION 2. We argue these cases one by one. In each of these settings, fixing i , v_i , v'_i and o_i , we explicitly construct a common upper bound of v_i and v'_i with respect to o_i , as described in Definition 5.

- **Common outcome, complete domain.** Let v''_i be such that $v''_i(o_i) = \max\{v_i(o_i), v'_i(o_i)\}$ and $v''_i(o'_i) = 0$ for $o'_i \in \mathcal{O}_i \setminus \{o_i\}$.
- **Single-parameter agents.** If $o_i = 0$, then let $v''_i(1) = 0$; otherwise, let $v''_i(1) = \max\{v_i(1), v'_i(1)\}$.
- **Combinatorial auctions, single-minded agents.** Let v_i be described by S_i and x_i , and v'_i by S'_i and x'_i . Let v''_i be induced by S''_i and x''_i , where $S''_i = o_i$, and $x''_i = \max\{x_i, x'_i\}$. Consider v_i , and observe that for any $o'_i \subseteq M$, if $o_i \not\subseteq o'_i$, then

$$v''_i(o_i) - v''_i(o'_i) = v''_i(o_i) \geq v_i(o_i) \geq v_i(o_i) - v_i(o'_i).$$

If $o_i \subseteq o'_i$, then because v_i is monotone,

$$v''_i(o_i) - v''_i(o'_i) = 0 \geq v_i(o_i) - v_i(o'_i).$$

Similarly, for v'_i , we always have

$$v''_i(o_i) - v''_i(o'_i) \geq v'_i(o_i) - v'_i(o'_i).$$

- **Combinatorial auctions, valuations “between” additive and XOS.** Recall that an XOS valuation is the pointwise maximum of a number of additive valuations, each of which is a clause. Let c_i be the clause of v_i such that $v_i(o_i) = c_i(o_i)$, and c'_i be the clause of v'_i such that $v'_i(o_i) = c'_i(o_i)$. Note that c_i (resp. c'_i) also satisfies for any set of items $T \subseteq M$, $c_i(T) \leq v_i(T)$ (resp. $c'_i(T) \leq v'_i(T)$). Let v''_i be an additive valuation such that for each item j ,

$$v''_i(\{j\}) = \begin{cases} \max\{c_i(\{j\}), c'_i(\{j\})\}, & \text{if } j \in o_i \\ 0, & \text{otherwise} \end{cases}.$$

Clearly $v''_i \in \mathcal{V}_i$, because it is additive. For any $o'_i \subseteq M$,

$$\begin{aligned} v''_i(o_i) - v''_i(o'_i) &= \sum_{j \in o_i \setminus o'_i} v''_i(\{j\}) - \sum_{j \in o'_i \setminus o_i} v''_i(\{j\}) \\ &= \sum_{j \in o_i \setminus o'_i} v''_i(\{j\}) \geq \sum_{j \in o_i \setminus o'_i} c_i(\{j\}) && \text{(construction of } v''_i) \\ &\geq \sum_{j \in o_i \setminus o'_i} c_i(\{j\}) - \sum_{j \in o'_i \setminus o_i} c_i(\{j\}) \\ &= c_i(o_i) - c_i(o'_i) = v_i(o_i) - c_i(o'_i) && \text{(property of } c_i) \\ &\geq v_i(o_i) - v_i(o'_i). && \text{(property of } c_i) \end{aligned}$$

Similarly,

$$v''_i(o_i) - v''_i(o'_i) \geq v'_i(o_i) - v'_i(o'_i).$$

- **Combinatorial auctions, valuations “beyond” subadditive.** Let $C = \max\{v_i(M), v'_i(M)\}$, and v''_i be such that $v''_i(\emptyset) = 0$, $v''_i(T) = 2C$ if $o_i \subseteq T$, and $v''_i(T) = C$ otherwise. It is easy to check v''_i is subadditive, and so $v''_i \in \mathcal{V}_i$. For any o'_i , if $o_i \subseteq o'_i$, because v_i is monotone,

$$v''_i(o_i) - v''_i(o'_i) = 0 \geq v_i(o_i) - v_i(o'_i).$$

If $o_i \not\subseteq o'_i$, then

$$v''_i(o_i) - v''_i(o'_i) = C \geq v_i(M) \geq v_i(o_i) - v_i(o'_i).$$

Similarly,

$$v''_i(o_i) - v''_i(o'_i) \geq v'_i(o_i) - v'_i(o'_i).$$

- **Metric space.** Let v_i be induced by x_i and v'_i by x'_i . Let v''_i be induced by $o_i \in \mathcal{O} \subseteq X_i$. Then for any $o'_i \in \mathcal{O} \subseteq X_i$,

$$v''_i(o_i) - v''_i(o'_i) = d_i(o_i, o'_i) \geq d_i(x_i, o'_i) - d_i(x_i, o_i) = v_i(o_i) - v_i(o'_i).$$

And similarly

$$v''_i(o_i) - v''_i(o'_i) \geq v'_i(o_i) - v'_i(o'_i). \quad \square$$

C.3 Proof of Theorem 3

PROOF OF THEOREM 3. Fix an IC and NB mechanism $M = (f, p)$. By Proposition 2, the payment characterization (Theorem 2) applies in this setting. To this end, for each $o_c \in \mathcal{O}_1 = \dots = \mathcal{O}_n$, let $p_i(o_c)$ be agent i 's payment when the common outcome is o_c . For each valuation vector v , consider the following transformed valuation vector \hat{v} , where for each agent i and outcome o_c ,

$$\hat{v}_i(o_c) = v_i(o_c) - p_i(o_c).$$

Observe that for any i and $v_i \in \mathcal{V}_i$, letting $<_i^{\text{int}} = <_i^{\text{int}}(v_i)$, for any $o_c, o'_c \in \mathcal{O}_1 = \dots = \mathcal{O}_n$,

$$(o_c, p_i(o_c)) <_i^{\text{int}}(o'_c, p_i(o'_c)) \iff \hat{v}_i(o_c) > \hat{v}_i(o'_c) \vee (\hat{v}_i(o_c) = \hat{v}_i(o'_c) \wedge o_c <_i^{\text{lb}} o'_c).$$

Now consider the behavior of \hat{f} , the transformed version of f , over all possible transformed valuations $\hat{\mathcal{V}} = \{\hat{v} \mid v \in \mathcal{V}\}$, where we define $\hat{f}(\hat{v}) = f(v)$. Then, the above observation gives a way to derive $<_i^{\text{int}}$ from \hat{v} directly, and so we can define IC for \hat{f} based on $<_i^{\text{int}}$. Note that \hat{f} corresponds bijectively to f , and $(\hat{f}, p^{\text{zero}})$ is IC iff (f, p) is IC, where p^{zero} assigns 0 payment to all agents in all cases. In other words, one can view \hat{f} as an IC mechanism over $\hat{\mathcal{V}}$ without payments. Suppose $|\mathcal{O}_1| = |\mathcal{O}_2| = \dots = |\mathcal{O}_n| = k$. Observe that such transformed valuations $\hat{\mathcal{V}}$ can induce all possible strict total orders over outcomes: for example, in order for agent i 's internal component of preference to be

$$o_c^{(1)} <_i^{\text{int}} o_c^{(2)} <_i^{\text{int}} \dots <_i^{\text{int}} o_c^{(k)},$$

we only need

$$\hat{v}_i(o_c^{(1)}) > \hat{v}_i(o_c^{(2)}) > \dots > \hat{v}_i(o_c^{(k)}).$$

Such a transformed valuation can be obtained from, for example, $v_i \in \mathcal{V}_i$ such that

$$v_i = (k - i) + p_i(o_c^{(i)}).$$

Now since \hat{f} is IC, it must induce a social choice rule over all strict total orders over the outcomes (i.e., it cannot assign different outcomes to two transformed valuation vectors inducing the same orders for all agents simultaneously), and moreover, this social choice rule cannot be manipulable. By the Gibbard-Satterthwaite Theorem [Gibbard, 1973, Satterthwaite, 1975], this social choice rule must either be dictatorship or use only two outcomes. The same characterization immediately applies to \hat{f} , and by the payment characterization extends to (f, p) , which concludes the proof. \square

C.4 Proof of Theorem 4

PROOF OF THEOREM 4. For any IC, IR and NB mechanism $M = (f, p)$, let $o^{(1)}, \dots, o^{(n+1)}$ be a decision list implementing M (with payment vectors $p(o^{(1)}), \dots, p(o^{(n+1)})$ and exception lists $E(o^{(1)}), \dots, E(o^{(n+1)})$). Such a decision list exists by Theorem 5 (we do not even require this list to be order-oblivious). Without loss of generality, assume each agent can sometimes get the item – otherwise we can remove the agents who never get the item, since they can never affect the outcome and payments because of NB. Note that the empty outcome $(0, \dots, 0)$ can never appear in an exception list of another outcome $o^{(j)}$, because in that case $o^{(j)}$ can never happen (since the payment vector corresponding to the empty outcome must be $(0, \dots, 0)$, and the empty outcome is always satisfied). Moreover, if the empty outcome is ordered before another outcome $o^{(j)}$, $o^{(j)}$ must be in the exception list of the empty outcome, because, again, otherwise $o^{(j)}$ can never happen. Given these facts, we can without loss of generality move the empty outcome to the end of the list, and assume $o^{(n+1)} = (0, \dots, 0)$. Moreover, we renumber the agents so that $o_i^{(i)} = 1$ for all $i \in [n]$. That is, in the i -th outcome, agent i receives the item. For brevity, let $q_i = p_i(o^{(i)})$ for each $i \in [n]$.

Now observe that for any two agents i and i' , it cannot be the case that $o^{(i)} \in E(o^{(i')})$ and $o^{(i')} \in E(o^{(i)})$ simultaneously. This is because if that happens, then for the valuation vector v where $v_i = q_i$, $v_{i'} = q_{i'}$, and $v_{i''} = 0$ for $i'' \notin \{i, i'\}$, we must have $f(v) \notin \{o^{(i)}, o^{(i')}\}$. However, for the valuation vector v' where $v'_i = q_i$ and $v'_{i''} = 0$ for $i'' \neq i$, we must have $f(v') = o^{(i)}$.⁹ In other words, i' can affect i 's outcome and payment without affecting i'' 's own outcome or payment, which violates NB.

We define the following “domination” binary relation over agents: we say i dominates i' if $o^{(i)} \in E(o^{(i')})$, or $i < i'$ and $o^{(i')} \notin E(o^{(i)})$. Intuitively, if i dominates i' , then if $o^{(i)}$ is feasible (i.e., $v_i \geq q_i$) then $o^{(i')}$ cannot be chosen. Observe that any two different agents i and i' are comparable under the domination relation, i.e., one of the two must dominate the other, and it is never the case that the two agents dominate each other simultaneously (because it cannot be the case that $o^{(i)} \in E(o^{(i')})$ and $o^{(i')} \in E(o^{(i)})$ simultaneously, as argued above).

Below we argue by contradiction that the domination relation actually is a total order over agents. In particular, it does not have 3-cycles. Suppose otherwise, i.e., there exist i, i' and i'' , such that i dominates i' , i' dominates i'' , and i'' dominates i . Then for the valuation vector v where $v_i = q_i$, $v_{i'} = q_{i'}$, $v_{i''} = q_{i''}$ and $v_{i'''} = 0$ for all $i''' \notin \{i, i', i''\}$, we must have $f(v) \notin \{o^{(i)}, o^{(i')}, o^{(i'')}\}$. However, for v' where $v'_i = q_i$, $v'_{i'} = q_{i'}$ and $v'_{i'''} = 0$ for all $i''' \notin \{i, i'\}$, we must have $f(v') = o^{(i)}$. This means i'' can affect i 's outcome and payment without affecting i'' 's own outcome or payment, violating NB.

Now we argue that the total order given by the domination relation is precisely one possible order that can be used in a sequential posted-price mechanism to implement M . In particular, the following sequential posted-price mechanism is equivalent to the decision list implementing M : repeatedly visit the agent who dominates all other agents among unvisited ones. Upon visiting each agent i , offer the price q_i . If $v_i \geq q_i$, then let $f(v) = o^{(i)}$ and $p(v) = p(o^{(i)})$. Otherwise, continue the process. Terminate the process and let $(f(v), p(v)) = ((0, \dots, 0), (0, \dots, 0))$ if there is no unvisited agent left. One can check this sequential posted-price mechanism does precisely what the decision list does. \square

⁹One subtlety is it is possible that there exists some i'' where $q_{i''} = 0$. This does not affect the argument much, because such an i'' cannot appear in the exception list associated with any other nonempty outcome, and i'' must be agent n . Then one can assume without loss of generality that $i' \neq n$, and the rest of the argument still works. Similarly one can fix the argument for the nonexistence of 3-cycles below.

C.5 Proof of Lemma 1

PROOF OF LEMMA 1. Suppose to the contrary that both $f(v) = o$ and $f(v') = o'$. Choose v'' such that $v''_i = \min\{v_i, v'_i\}$ for each $i \in [n]$. The plan is to argue that $f(v'') = f(v)$ and $f(v'') = f(v')$ simultaneously, a contradiction. We only need to show $f(v'') = f(v) = o$, for which it suffices to inductively argue that for each $i \in [n]$,

$$f(v) = f(v_1, \dots, v_{i-1}, v''_i, v''_{i+1}, \dots, v''_n) = f(v_1, \dots, v_{i-1}, v_i, v''_{i+1}, \dots, v''_n).$$

Fix some $i \in [n]$. By the induction hypothesis,

$$f(v_1, \dots, v_{i-1}, v''_i, v''_{i+1}, \dots, v''_n) = f(v) = o.$$

Observe that $v_i \geq v''_i$ since v satisfies both o and o' , and v'' is the maximum of $p(o)$ and $p(o')$. Consider the following 2 cases:

- $o_i = 0$. In this case, because M is IC, f must be monotone (by Myerson's characterization of single-parameter IC mechanisms [Myerson, 1981]), so decreasing agent i 's value from v_i to v''_i cannot change the personal outcome (or payment) of agent i , i.e.,

$$f_i(v_1, \dots, v_{i-1}, v''_i, v''_{i+1}, \dots, v''_n) = f_i(v_1, \dots, v_{i-1}, v_i, v''_{i+1}, \dots, v''_n).$$

Now since M is NB, the above implies

$$f(v_1, \dots, v_{i-1}, v''_i, v''_{i+1}, \dots, v''_n) = f(v_1, \dots, v_{i-1}, v_i, v''_{i+1}, \dots, v''_n).$$

- $o_i = 1$. In this case, because M is IC, fixing other agents' valuations, the menu agent i faces is a take-it-or-leave-it offer with price $p_i^o \leq v''_i$ (again by Myerson's characterization). So, decreasing i 's value from v_i to v''_i results in the same outcome and payment, because both v_i and v''_i are no smaller than $p_i(o)$. Again, since M is NB, this implies

$$f(v_1, \dots, v_{i-1}, v''_i, v''_{i+1}, \dots, v''_n) = f(v_1, \dots, v_{i-1}, v_i, v''_{i+1}, \dots, v''_n).$$

This finishes the induction step, and shows that $f(v) = f(v'')$. But then similarly one can show $f(v') = f(v'')$, which leads to a contradiction because $f(v) = o \neq o' = f(v')$. This concludes the proof. \square

C.6 Proof of Theorem 5

PROOF OF THEOREM 5. Given an IC, IR, and NB mechanism $M = (f, p)$, we explicitly construct a decision list implementing this mechanism. We first construct a "decision tree" (which is quite different from a normal decision tree), which we will later turn into a decision list. The decision tree we construct is of the following form: each node is associated with a set of outcomes, which we call the domain of the node, and a single outcome in the domain. In particular, the domain of the root is \mathcal{O} . Each node may have any number of children, with the constraint that the domain of any child of a node is a strict subset of the domain of its parent (so the decision tree must be finite). The way we choose an outcome using such a decision tree is the following: starting from the root, at every node, we check whether the valuation vector v satisfies the outcome associated with the node. If yes, then that outcome is the one we choose. Otherwise, we move on to an arbitrary child of the current node satisfying the following condition: v does not satisfy any outcome that is not in the domain of that child (we will see from the construction that such a child always exists), and repeat the above procedure. The intuition behind the procedure will become clear momentarily.

Now we construct the tree. For any set of outcomes $S \subseteq \mathcal{O}$, let v^S be such that for each $i \in [n]$,

$$v^S_i = \max\{p_i(o) \mid o \in S\}.$$

We start from the root, whose domain is \mathcal{O} , and let $f(v^{\mathcal{O}})$ be the outcome associated with the root. This means the tree should choose $f(v^{\mathcal{O}})$ whenever it is satisfied. This is in fact consistent with

f , because for any other outcome $o \in \mathcal{O}$, $v^{\mathcal{O}}$ satisfies both $f(v^{\mathcal{O}})$ and o . By Lemma 1, whenever $f(v^{\mathcal{O}})$ is satisfied by a valuation vector v , $f(v) \neq o$. This holds for any $o \in \mathcal{O} \setminus \{f(v^{\mathcal{O}})\}$, so it must be the case that $f(v) = f(v^{\mathcal{O}})$ whenever v satisfies $f(v^{\mathcal{O}})$.

Now we describe the way we construct the children of a node (e.g., the root), which, applied recursively, gives the entire tree. Let D be the domain of the parent node, and o be the outcome associated with the parent node. Each child corresponds to a maximal subset D' of D such that $v^{D'}$ does not satisfy o . Note that there may be many such maximal subsets, and we construct a child for each of these maximal subsets. For the child whose domain is D' , again, we let the outcome associated with that child be $f(v^{D'})$. Observe that when the tree chooses $f(v^{D'})$ at this node, this choice must also be consistent with f . This is because if we reach this node, then it must be the case that the valuation vector v can only satisfy outcomes in D' . And again by Lemma 1, if v satisfies $f(v^{D'})$, then it must be the case that $f(v) = f(v^{D'})$.

Now consider the correctness of the decision tree constructed, i.e., whether it implements f . Above we have argued that whenever the decision tree chooses an outcome, that outcome must be consistent with the choice by f . So we only need to show the decision tree in fact always outputs an outcome. Consider any valuation vector v . First observe that v satisfies at least 1 outcome, i.e., the empty outcome $(0, \dots, 0)$. Below we argue that at any node with domain D and associated outcome o , there is always a child that we can go to, if v does not satisfy o . In fact, let $S \neq \emptyset$ be the set of outcomes that v satisfies. Because of the way we go down the tree, we must have $S \subseteq D$. Also, we have $v_i \geq v_i^S$ for each $i \in [n]$, simply by the definition of S . If v does not satisfy o , then clearly v^S does not satisfy o either. So S must be a subset of the domain of at least 1 child of the current node, because we choose the domains of the children in a maximal way. That child is a node we can go to. Now because the tree is finite, we cannot keep going down forever, so we must be able to choose an outcome somewhere in the tree. Together with all the arguments above, this shows that the tree we construct in fact implements f (and M because $p(v) = p(f(v))$).

Now we turn the decision tree into a decision list. First observe that when evaluating the decision tree, the order in which we visit the nodes does not matter (although the specific way of evaluating the tree discussed above is instrumental in proving the correctness of the construction). More specifically, let $\{(D^{(j)}, o^{(j)})\}_{j \in [K]}$ be the collection of domain-outcome pairs of all nodes of the decision tree, where K is the number of nodes. To evaluate the decision tree, it suffices to check $(D^{(j)}, o^{(j)})$ one by one in any order, and choose the first $o^{(j)}$ where v satisfies $o^{(j)}$ and does not satisfy any outcome not in $D^{(j)}$. In fact, since the decision tree implements f , exactly one of these pairs will satisfy the condition of being chosen. Now we merge pairs with the same outcome: for any outcome $o \in \mathcal{O}$, let

$$D(o) = \bigcup_{j: o^{(j)}=o} D^{(j)}.$$

Consider the new collection of pairs $\{(D(o), o)\}_{o \in \mathcal{O}}$, and observe that it is equivalent to the old list: when the old list chooses some outcome, the new list chooses the same outcome. One subtlety is that the new list may choose multiple outcomes simultaneously, which would make it ill-defined. However, this cannot happen because whenever the new list chooses an outcome, it must be the outcome assigned by f . To see why this is the case, consider any valuation vector v . Let $(D(o), o)$ be a pair in the new list such that v satisfies o and does not satisfy any outcome not in $D(o)$. By Lemma 1, because v satisfies o , for any $o' \in D(o) \setminus \{o\}$ that is also satisfied by v , $f(v) \neq o'$. On the other hand, since M is IR, $f(v)$ must be satisfied by v , so the only option left is $f(v) = o$. Now let the exception list of each outcome o be $E(o) = \mathcal{O} \setminus D(o)$. This gives an order-oblivious decision list as stated in Theorem 5. \square