Informational Robustness in Mechanism Design A Case for (and Against) the Worst Case*

Benjamin Brooks Songzi Du

July 16, 2025

Abstract

A central issue for mechanism design is how to identify theoretical environments that will lead to useful insights about optimal mechanisms. One desideratum is that the agents' private information structure should be relatively *simple*, so that the corresponding optimal mechanisms are also not overly complex. Another important criterion is that the resulting optimal mechanisms are *portable*, meaning that they will continue to perform well in environments other than the one they were optimized for. We argue that by focusing on environments that are the most challenging for designer, the informationally-robust approach will tend to identify information structures and mechanisms that are both simple and portable. We survey a recent literature that operationalizes this idea to derive novel mechanisms for a variety of mechanism design problems. We place this literature in the broader context of robust mechanism design, and we also discuss weaknesses and shortcomings of the informationally-robust approach.

KEYWORDS: Mechanism design, information design, dual reduction, max-min, Bayes correlated equilibrium, robustness.

JEL CLASSIFICATION: C72, D44, D82, D83.

*Brooks: Department of Economics, University of Chicago, babrooks@uchicago.edu; Du: Department of Economics, University of California San Diego, sodu@ucsd.edu. We are grateful for numerous conversations and collaborations with Dirk Bergemann, Joey Feffer, Alexander Haberman, Lars Hansen, Emir Kamenica, Stephen Morris, Roger Myerson, Philip Reny, and Linchen Zhang. This research has been supported by NSF #2215259 and #2215475.

1 A Case for the Worst Case

1.1 Finding the Right Model

Roger Myerson has told the following story. In the late 1970s, he was working on what eventually became his paper on revenue maximization in auctions. He had realized that the revelation principle could be used to reduce the problem to a linear program, with parameters specifying the distribution of values. But the question remained: What do optimal auctions look like, and how do they depend on the value distribution? To get some inspiration, Roger took a computational approach. At that time, Northwestern had a communal mainframe that could run the simplex algorithm. So, at Roger's request, the school gave him a $100 budget to buy computer time for his research on optimal auctions. He decided to try out a few minimalist examples, consisting of two buyers and two possible values. He coded up three instances of the linear program (on punch cards) and ran them through the number cruncher, for 25 cents a pop. For the first two simulations, an astonishing result was that private information seemed to not matter at all for revenue; the seller was able to do just as well as if they knew the buyers' values, and there were no private information. But the third simulation was more challenging for the seller; the presence of private information was a substantive constraint on how much revenue could be generated. Of course, in the first two programs, the distribution was such that values were correlated, and in the third they were independent. As they say, the rest is history. Roger decided that for his theory, he should develop a general analysis when the bidders have independent types. This led to a number of insights that hugely influenced the subsequent development of mechanism design and auction theory, including the revenue equivalence theorem and the celebrated result that first- and second-price auctions with reserve prices can be rationalized as optimal auctions. A correlated example appeared at the end of Roger's paper, illustrating cases where the theory seemed to lead to a "very strange auction" with peculiar "side bets," with the seller sometimes paying money to the buyers (Myerson, 1981). The rest of the $100 budget was left unspent.

A critical issue that this anecdote highlights is the following: Some models of private information and preferences seem more useful than others for gaining insight into efficacious designs for mechanisms. To make progress in mechanism design, we have to somehow identify the parametric assumptions that generate useful results, while also reconciling the theory with those cases where the implications are less compelling. That issue, generally speaking, is the subject of this article. In the anecdote, the specific context was revenue maximization in private-value auctions. But the issue is much broader and affects all of mechanism design. For the most part, our discipline follows a relatively ad hoc approach of specifying particular functional forms for preferences and private information, often motivated by institutional details. In the formal treatment below, these objects are modeled jointly as an *information structure*. In spite of the nominal emphasis on information, an information structure is actually a description of both the possible ex post preferences the agents might have as well as their private information thereabout. Given a fixed information structure, we can analyze mechanisms that are optimal for that environment.

An obvious concern with this approach is that the space of possible information structures is large and diverse, and an ad hoc approach to exploring that space may miss the information structures that would be the most insightful. As an alternative, a recent literature has proposed a more systematic approach for sifting through the myriad possible environments: We first fix certain payoff relevant aspects of the environment that we think the designer has a good handle on. Subject to those constraints on fundamentals, we then focus our attention on the forms of private information that are the most challenging for the designer. Thus, the agents' private information about preferences is endogenously determined as the one that provides the greatest opposition to the designer's goals. The hope is that this approach will identify environments that will be especially insightful. Over the course of this article, we will review that literature and what it has achieved so far. We will also explore higher level reasons for why that approach may or may not succeed in its aims. The rest of this introduction lays out the main ideas informally, with a mathematical treatment to follow.

1.2 The Problem in Mechanism Design

We will later consider general purpose mechanism design, but for now, the optimal auctions problem provides a useful and familiar setting in which to develop our central thesis. Indeed, the independent private value (IPV) model has had such enduring significance for mechanism design and auction theory, going all the way back to Vickrey (1961), that it seems fair to hold it up as the gold standard for an information structure that economists and game theorists consider to be a useful benchmark. This may be due to a combination of factors, one of which is that the information is expressed in a seemingly natural language, where the units are just dollars of willingness to pay. This is relatively easy to map into the human experience, where we regularly compare the value derived from consumption of a good to its price.

At the same time, we see no reason to favor *independence* per se on grounds of realism. If anything, it would seem like an extraordinary coincidence if the values just happened to be *exactly* independent. This point has been emphasized by the literature on the full surplus extraction paradox (Myerson, 1981; Crémer and McLean, 1985, 1988; McAfee et al., 1989; McAfee and Reny, 1992). And yet, the independent and correlated models differ radically in the kind of predictions they make for optimal mechanisms and in how they have been received by economists. In the independent case, the optimal mechanisms seem to have properties that make them desirable in other non-IPV environments, such as affiliated values (Engelbrecht-Wiggans, Milgrom, and Weber, 1983; Milgrom and Weber, 1982) and beyond (Chung and Ely, 2007; Bergemann, Brooks, and Morris, 2017, 2019), even if they are not exactly optimal for those settings. This has engendered a view that the IPV model distills fundamental insights about behavior and institutional design that are relevant in a wide variety of settings. By contrast, the mechanisms rationalized by correlated private values seem to be much less portable to other environments. The question of how to think about these strange auctions, and their broader implications for mechanism design, has confounded game theorists for forty-five years.

So why should the independent case produce natural looking mechanisms that seem to have a life beyond the model, while the correlated case leads to designs for mechanisms that seem bizarre and implausible? Both of these models make heavy use of quasilinear expected utility, both assume that the agents make sophisticated probabilistic assessments of how others will behave (consistent with a common prior), and in both cases Bayes Nash equilibrium is employed as the solution concept. One could argue that the payments to the bidders or extraordinarily large payments to the seller that arise with correlated values would be difficult to implement in practice. But even with constraints on the sign or size of payments, "bets" about others' information would still arise in optimal mechanisms.

In our view, a key issue is not with correlation per se. Rather, the problem is that, placing ourselves in the shoes of the designer, we do not know *which* form of correlation is empirically relevant, meaning the precise probabilistic assessments that the agents maintain over one another's private information. Obviously, there is no general correct answer, and the exact probabilities would inevitably depend on the particular setting. But if we view a mechanism as a long-lived institution, that is meant to perform well under varied circumstances and as the situation changes over time, then we should look for it to work well under lots of different correlation structures.

Now, it seems that there is an easy solution. Why model just one form of correlation? Why not embrace the fact that there may be different correlation structures, by simply modeling a larger information structure, that includes all of these possible worlds within it? And on the mechanism side, we can make that richer as well. Have more actions for the agents, corresponding to different possibilities for the structure of correlation. Let the mechanism discover "online" which is the true form of correlation.[1] In fact, why not go even further by allowing for the possibility that agents learn their respective values, and then observe even more information about others' values? Chung and Ely (2007) have gone so far as to model mechanism design on the entire universal type space, the infinite dimensional vector space consisting of all sequences of higher-order beliefs![2] And why not go even further than that, and add in all of the possible correlation devices and conditional information that is suppressed by the belief hierarchies (Ely and Peski, 2006; Liu, 2009)?

At this point, we must interject and remark on an important interpretational distinction between the model of the mechanism on the one hand, and the model of the agents' preferences and information on the other hand. In many real-world settings, the mechanism is not just a metaphor for a naturally occurring market structure; it is intentionally an constructed institution. Indeed, auctions and other designed markets are ubiquitous in our economy. Moreover, there is an abundance of historical instances where the abstract game-theoretic formalism of a mechanism has been translated into a functioning real-world institution. Thus, from a normative perspective, the abstract language and rules of a mechanism represent a genuine practical possibility.

[1]This rhetorical point has been made by Börgers (2017), where it is framed as a critique of the worst-case analysis surveyed in this article. See also Kambhampati (2025) for a distinct approach that considers lexicographic preferences over possible environments.

[2]Chung and Ely (2007) consider a subset of the universal type space of Mertens and Zamir (1985) in which each agent knows their own value. Of course, we might wish to go even more general and allow for interdependence!

The information structure is, by contrast, a much looser metaphor for reality. Of course, human beings possess private information, but that information is in reality a messy, complicated object, that cannot be concisely expressed through natural language. For example, in real estate, prospective homes have a huge number of characteristics, and individuals vary in their needs and desires and constraints. An agent's knows all of the listing details, tours, market research, introspection, and much much more information that they exposed to over the course of a home search. A mathematically tractable model of information is a dramatically simplified artifice, useful for analysis but reflecting reality only at an atmospheric level of abstraction. A further issue is that even if we are willing to embrace the information structure as a descriptive theory, it may only be valid as an implicit *representation* of the agents' choices. We should not take the abstract language of signals in an information structure literally as something that agents are consciously aware of or could articulate through interaction with a mechanism designed by mathematical economists.

Returning to the auction problem, suppose we take the information structure to be so rich that it covers all of the possible correlation structures. Then an optimal mechanism will so too build in all kinds of elaborate ways for the agents to behave, corresponding to different optimal mechanisms for different submodels of information. At the very least, this would consist of all of the possible lotteries, as mappings from others' information to dollar amounts, needed to separate agents by their beliefs. If we wished to go even further and build a "grand" mechanism to cover all environments, then the language would need to be as rich, abstract, and artificial as that used to describe information structures themselves. Do we actually believe that human beings would be able to play such an elaborate mechanism as a designer intends? The burden on the agents to coordinate their play as intended for "their" information structure would be absolutely mind-boggling. Our point is that if the model of information is too complex, then the optimal mechanism will be overly complex as well, so that equilibrium is no longer a compelling prediction for behavior, and the connection between theory and reality will be lost.

1.3 What are we after?

All of this indicates that if we wish to obtain optimal mechanisms that have a hope of actually being implemented in a manner that resembles the theory, the information has to have a relatively simple structure. At the same time, the hope would be to use the theory of optimal mechanisms to discover designs that have a life beyond the particular model of information that gave rise to them. At the very least, we would hope that performance of the mechanism would not degrade precipitously if the environment turns out to be different from our baseline assumptions. Putting it a bit differently, each information structure has its own lesson about the right way to align the the agents' incentives with the designers' goals, and these lessons are incorporated into the corresponding optimal mechanism. In our view, a desirable feature of a model of information is that its associated lesson remains relevant, even if the true model of information turns out to be different.

Framed in this manner, the IPV model seems like a grand success: Each agent only possesses information about their own value, and not about the values or information of others. Moreover, this information can be "ordered" in a natural way according to the values. This orderly structure is reflected in the optimal mechanisms as well, in the manner in which one-dimensional bids in a first- or second-price auction correspond to increasing allocation and payments. Moreover, the "lesson" from the IPV model about how to structure incentives seems relevant well beyond that environment. In particular, an agent's value is a measure of the strength of their objections to what the designer wants to achieve (i.e., rent extraction), where by "objection" we mean refusing to participate in the mechanism, because the resulting outcome would be too deleterious. But not only can agents object to the mechanism: they can also behave *as if* their objections are stronger than they truly are. A particularly salient special case is pretending that one's objections are just slightly stronger than they are in actuality (e.g., behaving as if one's value is slightly lower than the true quantity). Optimal mechanisms have to manage the rents that agents obtain from a combination of being able to object and being able to overrepresent their objections. In the specific case of IPV with regular distributions, rents are optimally managed by linking higher priority in the allocation to higher payments. While this simple structure is exactly optimal only in certain settings, as we mentioned previously, there are numerous results in the literature arguing for the efficacy of such mechanisms in non-IPV environments.

By contrast, if we limit our scope to a particular instance of the correlated values model (as we must to keep the optimal auction manageable in size and complexity), then the "lesson" is that given heterogeneous probabilistic assessments, the designer can forestall the agents' from overrepresenting their objections by linking each action to a bespoke wager about others' information, where each of these wagers would appear extremely unfavorable except for a given belief and value. But because the mechanism is tailored to the specific probabilistic assessments that are hypothesized, they lack portability to other environments.

1.4 Simplicity and Portability

Thus, in the context of optimal auctions with private values, the independent model seems to lie at the sweet spot of being simple enough that optimal mechanisms are of a manageable complexity, while also generating portable insights about how to manage incentives. Now, one might reasonably object to our use of the word "simple" in such a vague and ill-defined manner. The truth is that we do not know how to define or quantify "simplicity" in a general manner. This inability represents a significant methodological obstacle in game theory and mechanism design, which we will return to at the end of this article. In lieu of a definition, we will instead take it as a premise that the regular IPV model is simple. Indeed, there are numerous examples of papers that have tried to generalize the regular IPV model by identifying sufficient conditions on the information structure so that the same pattern of equilibrium constraints characterizes the optimal mechanism: Roughly speaking, the condition is or should imply that private information can be ordered by the strength of the opposition to the designer's goals, and that the critical equilibrium constraints are that agents not be tempted to slightly overrepresent their objections.[3] We refer to such information structures as *regular*, and we take it as an axiom that regularity is a sufficient condition for an information structures to be considered "simple."[4]

Now, if private-value auctions were the only problem we were interested in, then perhaps we would declare victory with IPV and call it a day. But mechanism design is much broader than that. And it is not unfair to say that for most problems of interest, ranging from multiproduct monopoly, bilateral trade, public goods, and partnership dissolution, there is no set of assumptions about preferences and information that have achieved the status of a commonly accepted benchmark, like the IPV model in private-value auctions. Even within auctions, there are natural alternative payoff environments, such as those with interdependence and correlation in values, for which there is no satisfying theory of optimal mechanisms. For example, the mineral rights model of Wilson (1977) strikes many as a natural metaphor for noisy information about a pure common value, but because the signals are richly correlated, "strange auctions" like those described by Myerson (1981) are again optimal and completely neutralize private information (McAfee, McMillan, and Reny, 1989).

So how do we emulate or improve on the success of IPV in these other domains? And how do we build productive theories of optimal mechanisms? Going back to the opening anecdote, we do not see it as a coincidence that the IPV model was both the most challenging for the designer, among those models considered, and also has proven the most insightful. Indeed, this brings us to our central thesis: the informational environments that are most challenging for the designer will tend to lead us to practically useful insights about mechanism design. Why? Well, it has to do with the twin goals of *simplicity* and *portability*.

To explain our views on simplicity, it is helpful to introduce some concepts. Let us define the *potential* of an information structure to be the optimal value of the designer, i.e., the designer's highest payoff across all mechanism and equilibria. We take the potential as the measure of how challenging the environment is. When values are richly correlated, the agents know much more than the designer about one another's values and information. However, this rich information is actually quite beneficial to the designer, because it allows the designer to construct wagers about others' information that represent very different lotteries to different agents. So, the agents having very complex information may actually be beneficial to the designer, and lead to a high potential. And yet, it is certainly not the case that more information for the agents always raises the potential. Indeed, consider the case of a single buyer: if they knew nothing about their value (except for its prior distribution), then it would be possible to extract all of the surplus, simply by offering the good at a posted price equal to the ex ante expected value! Whether more private information leads to higher or lower potential depends on what form it takes. The cases that are more challenging for the designer are when private information takes a disciplined form, that serves to inform the agents of when their objections to the mechanism are the strongest, and then amplify the constraints on the designer through agents' ability

[3] See, e.g., Bulow and Klemperer (1996), Chung and Ely (2007), Pavan, Segal, and Toikka (2014), Chen and Li (2018), and Yang (2024).

[4] We recognize that this is a controversial premise, and we will critique it below in Section 5.

to overrepresent their objections. Thus, low (or even minimum) potential seems like a criterion that will select for relatively simple forms for information.

To discuss portability, we introduce another concept: The *guarantee* of a mechanism is the lowest possible value to the designer, across all information structures and equilibria. The guarantee of a mechanism also isolates a "most challenging" environment, but it is the most challenging for that particular mechanism. By definition, a mechanism with a high guarantee must have structure that ensures at least that payoff for the designer, regardless of the environment. In that sense, mechanisms with higher guarantees incorporate within them more portable methods for aligning the agents' incentives with the designer's objectives.

So, we have two ideas: Information that minimizes the potential will tend to be simple, and represent the most efficient levering up of agents' objections, in order to constrain the designer as much as possible. And mechanisms that maximize the guarantee embody, in a certain sense, the most portable methods for aligning incentives with the designer's notion of welfare. How do these two come together?

We may dualize the mechanism designer's self into one half that is searching for the environment that is most challenging, meaning information structures with low potential, and the other half that is searching for portable mechanisms, meaning those with a favorable guarantee. This is a sort of zero-sum game that the designer plays against themself, to systematically and simultaneously explore both sides of a mechanism design problem. Of course, given an information structure and a mechanism, there is still the matter of which equilibrium will be played. But setting that aside for the moment, and pursing the analogy with zero-sum games, a very strong form of solution of this introspective game would be a *saddle point*: a potential-minimizing information structure and a guarantee maximizing mechanism, where the minimum potential is equal to the maximum guarantee. Such a saddle point would represent a confluence of the ideas we have espoused on simplicity and portability: The guarantee-maximizing mechanism, meaning the one with the most "portable" approach to aligning incentives, would also be an optimal mechanism for the most "challenging" information structure, so that if the latter turned out to have a relatively simple form for private information, then this would tend to be reflected in the mechanism as well.

1.5 A Brief Survey

The literature we survey below has shown that for many fundamental problems in mechanism design, such saddle points do exist. Moreover, whether or not a saddle point exists, there will always exist potential-minimizing information structures that resemble the IPV model in key ways. In particular, there will always exist a potential minimizing information that is regular, in the aforementioned generalized sense: private information can be ordered, in terms of the strength of objections to the designer's goals, and the key constraints that pin down the potential are that the agents must not want to behave as if their objections are slightly stronger than they truly are.[5] This structure is also reflected in the guarantee maximizing mechanisms: It is always possible to find a guarantee maximizer in which the agents' actions can be interpreted as an expression of the strength of their opposition to the designer's goals. The key issue in designing guarantee-maximizing mechanisms is trading off the value of the implemented outcome to the designer, versus the pressure it places on the agent to slightly *underrepresent* their objections. We will refer to a mechanism for which the guarantee is pinned down by equilibrium constraints of this form as also being *regular*, as a dual notion to that applied to information structures.

Thus, we may concisely express a key conclusion of this literature: there necessarily exist potential-minimizing information structures and guarantee-maximizing mechanisms that are regular. This is true at a remarkable level of generality, holding for any set of outcomes, any set of payoff-relevant states, and any expected utility preferences of the agents and the designer.[6]

A further key insight is that both of the programs of potential minimization and guarantee maximization can be formulated as *linear programs*, which lends analytical and computational tractability to the theory. In fact, there is sense in which these programs are approximate duals of one another, and so that solutions to the potential/guarantee problem can be viewed as saddle points in the sense of linear programming duality.

Finally, the literature has produced a number of fully worked applications to auctions, trading mechanisms, and the provision of public goods. These applications have produced novel designs for mechanisms in their respective settings.

The following sections survey these results in greater detail. We will describe the general theory and the main results outlined in the preceding paragraphs. We will also provide two worked examples: one involving a common value auction, and another involving bilateral trade. These examples will serve to illustrate how the theory can be operationalized to solve for guarantee-maximizing mechanisms and the potential-minimizing information structures that rationalize them. Towards the end of the article, we will take a broader view and cast these results within the broader literature on robust mechanism design. The article concludes with a critique of the informationally-robust approach, and a discussion of what we see as the most promising avenues for further development of the theory.

2 A Model of Mechanism Design

We now transition to a more formal analysis, beginning with a mathematical description of the joint mechanism and information design problem described in the introduction. The

[5]In the formal statements of these results, we describe approximate potential-minimizing information, and the regularity condition holds approximately as well. A similar caveat applies to our description of guarantee-maximizing mechanisms as regular.

[6]To economists working on mechanism design, this result may seem counterintuitive: "regularity" is usually conceived of as a property on primitives that *implies* the aforementioned structure on the critical equilibrium constraints. Instead, what we are saying is that if we *derive* the information structure and mechanism according to the min potential or max guarantee, then we will endogenously arrive at functional forms that are regular. Thus, the informationally robust analysis is a tool for identifying primitives which will imply an order on private information or actions, for which local equilibrium constraints are sufficient.

economy consists of a finite number of agents, whom we index by $i = 1, \ldots, N$, as well as a mechanism designer. There is some fundamental uncertainty which affects preferences, which is parametrized by a *state of the world*, denoted by $\theta \in \Theta$. The designer is in control of an *outcome* $\omega \in \Omega$. Both the agents and the designer have expected utility preferences over states and outcomes, which are represented by a utility index $u_i(\omega, \theta)$ for agent i, and a welfare function $w(\omega, \theta)$ for the designer.

The set of states (and other sets we introduce) may be finite or infinite, depending on the application, though for most of the formal results we describe in this survey, we will suppose that these sets (and other objects as well) are finite. We comment further on the finiteness assumption after presenting the rest of the model.

The designer can build a *mechanism* by which the outcome will be determined. Specifically, the mechanism consists of, for each agent, a set A_i of actions that can be taken in the mechanism, and a mapping m that associates to each profile of actions $a = (a_1, \ldots, a_N)$ a likelihood $m(\omega|a)$ that the outcome ω will be implemented. The entire mechanism is denoted by $M = (A, m)$, where $A = A_1 \times \cdots \times A_N$ is the set of action profiles. The mechanism should be thought of as an institution, such as an auction or a market exchange, erected by the designer. Let $\mathcal{M}$ be the set of finite mechanisms.

A subset of mechanisms that will be important in our theory are those that are *participation secure*, meaning that for every agent i there exists a *secure* action, which we label as $0 \in A_i$, and for which

$$\sum_{\omega} u_i(\omega, \theta) m(\omega|0, a_{-i}) \geqslant 0$$

for every θ and a_{-i}. Let $\mathcal{M}^0$ be the set of finite participation-secure mechanisms. A fundamental assumption of the theory is that a participation-secure mechanism exists.

The purpose of the mechanism is to incorporate private information that the agents have about θ into which outcome ω is implemented. This private information is described by an *information structure*, which consists a set of possible signals S_i for each agent i, and a joint distribution $\sigma \in \Delta(S \times \Theta)$, where $S = S_1 \times \cdots \times S_N$ is the set of signal profiles. The information structure is finite if S is finite.

An information structure is formally equivalent to a common prior type space in the sense of Harsanyi (1967), with one caveat: we have built into our definition the *common prior assumption*, that differences in beliefs are a result of differences in information. We will comment more on this assumption in Section 4.

For most of the theory, we hold fixed a prior distribution $\mu \in \Delta(\Theta)$. Thus, the uncertainty of the designer is not about payoff relevant fundamentals of the economy, but only about the agents' information. A natural extension of the exercise would be to also include uncertainty about μ. We discuss this possibility further in Section 4. But for now, we focus on the case where only information is unknown. Let $\mathcal{I}$ be the set of finite information structures for which the marginal of σ on θ is μ.

Given a finite mechanism and a finite information structure, a (behavioral) strategy for an agent associates to each signal s_i and action a_i a likelihood $b_i(a_i|s_i)$. A profile of strategies b can be identified with a mapping from signal profiles to lotteries over action profiles, where $b(a|s) = \prod_i b_i(a_i|s_i)$.

Given a finite mechanism M, finite information structure I, and strategy profile b, the ex ante expected utility for agent i is

$$U_i(M, I, b) = \sum_{a,s,\theta,\omega} u_i(\omega, \theta) m(\omega|a) b(a|s) \sigma(s, \theta),$$

and that of the designer is

$$W(M, I, b) = \sum_{a,s,\theta} w(\omega, \theta, \omega) m(\omega|a) b(a|s) \sigma(s, \theta).$$

Finally, b is a *(Bayes Nash) equilibrium* if $U_i(M, I, b)$ is greater than $U_i(M, I, (b'_i, b_{-i}))$ for every agent i and alternative strategy b'_i. Let $\mathcal{E}(M, I)$ denote the set of equilibria given a finite mechanism M and finite information structure I. We further denote by $\mathcal{E}^0(M, I)$ the subset of equilibria for which

$$\sum_{a,s_{-i},\theta,\omega} u_i(\omega, \theta) m(\omega|a) b(a|s) \sigma(s_i, s_{-i}, \theta) \geqslant 0$$

for all i and s_i, i.e., those equilibria that satisfy interim participation constraints.

In this theory, in order to know how a mechanism will perform, we need to specify both the information structure and the equilibrium being played. A key challenge for the designer is to determine which information structures will be useful for disciplining the design of the mechanism. As we argued in the introduction, environments that are especially challenging for the designer stand out as candidates that may guide us to especially portable designs for mechanisms. Here, "challenging" is quantified by the best performance that can be achieved by the designer, which we term the potential.

$$P(I) = \sup_{M \in \mathcal{M}} \sup_{b \in \mathcal{E}^0(M,I)} W(M, I, b). \tag{1}$$

The most challenging information structure is one that minimizes the potential. We denote the infimum potential across all finite information structures by

$$P^* = \inf_{I \in \mathcal{I}} P(I).$$

Relatedly, we seek mechanism that is portable across information structures. We quantify how portable a mechanism M is by its *guarantee*, defined as

$$G(M) = \inf_{I \in \mathcal{I}} \inf_{b \in \mathcal{E}(M,I)} W(M, I, b). \tag{2}$$

The most portable mechanism is one that maximizes the guarantee. We denote the highest possible guarantee over participation-secure mechanisms as

$$G^* = \sup_{M \in \mathcal{M}^0} G(M).$$

Note that we have modeled participation constraints in two different ways for the guarantee and the potential; for the latter we have used a standard interim participation constraint, while for the former we have used participation security. The latter implies the former: In particular, if M is participation secure, then for any I, any equilibrium of (M, I) must satisfy interim participation constraints; if not, then an agent could profitably deviate to the secure action.

The reason for adopting a different approach to participation constraints in the two problems is that individual rationality is not a property of a mechanism; it is a property of an equilibrium in a particular mechanism and information structure. But in formulating the guarantee-maximization problem, we wish to restrict attention to mechanisms in which agents are willing to participate, regardless of the information structure and equilibrium. In general, participation security is stronger than the requirement that every information structure has an equilibrium in which interim participation constraints are satisfied. We discuss this further with an example in Section 5.

An elementary observation is that $G(M) \leqslant P(I)$ for any participation-secure mechanism M and information structure I. The reason is as follows. First, The equilibrium therefore has an expected welfare for the designer that is less than $P(I)$. Hence, that $G(M)$ (the worst equilibrium outcome across all I') must be less than $P(I)$. It immediately follows that $G^* \leqslant P^*$. There are cases for which $G^* < P^*$, in which case we say there is a *duality gap*. The theory is especially powerful when there is no duality gap, so that the max guarantee is equal to the min potential. In this case, the potential-minimizing information structures certify that the guarantee-maximizing mechanisms are unimprovable, and vice versa.

An especially strong type of solution would be a *saddle point*: a pair $(\overline{M}, \overline{I})$ for which $G(\overline{M}) = P(\overline{I})$. For such a pair, $\overline{M}$ is necessarily a guarantee-maximizing mechanism (since no mechanism can have a guarantee higher than $P(\overline{I})$, and $\overline{I}$ is necessarily a potential-minimizing information structure (since no information structure can have a potential less than $G(\overline{M})$), and there is no duality gap. Moreover, we will show that saddle points possess additional mathematical structure that greatly facilitates their characterization.

Notice that we cannot appeal to a conventional minimax theorem for the existence of a saddle point. This is for a number of reasons: First, given a pair (M, I), we would consider the worst equilibrium for the designer when computing the guarantee, and the best equilibrium for the designer when computing the potential. Thus, this game between a guarantee maximizer and a potential minimizer is not actually zero-sum. Moreover, even if we fix an equilibrium selection rule, because the set of equilibrium outcomes can vary in a complicated way with the pair (M, I), we do not know of any linear or convex structure on equilibrium welfare that would allow us to apply the standard minimax theorems. Nonetheless, in the next section we will give intuitions for its existence and incentive structure, as well as examples in some concrete problems. Finally, with the restriction to finite mechanisms and information structures, the action spaces of the guarantee maximizer and potential minimizer may not be appropriately compact, in order for G^* and P^* to be attained exactly.

Now is a good moment to return to the issue of finiteness. In the definitions of a guarantee and a potential, we have restricted attention to finite mechanisms and finite information structures, following Brooks and Du (2024, 2025). For the general theory, this restriction is both conceptually beneficial and analytically convenient. On the conceptual side, if we allow for infinite mechanisms and information structures, then we would have to confront the possibility that equilibria do not exist. This is more than a technical concern;

if a mechanism's guarantee is only evaluated for those information structures in which an equilibrium exists, then we may reasonably worry that its guarantee holds only vacuously, because equilibria do not exist on those information structures which would actually be the most challenging. More subtly, even if equilibria exist, the set of equilibria may be unnaturally small due to a controversial exploitation of non-existence of best responses. This is related to critiques of results on full implementation using integer games that exploit non-existence of best responses. See Jackson (1992) and Abreu and Matsushima (1992b) for provocative discussion of these issues. But if the mechanism and information structure are both finite, then best responses always exist, and such concerns are moot. On the analytical side, as we will see, finiteness allows us to appeal to the powerful and elementary theory of finite dimensional linear programming for the main results.

Finiteness is a significant limitation, however, in that the supremum guarantee and infimum potential may only be achieved in the limit as the number of actions or signals goes to infinity. This is precisely what happens in the optimal auctions and bilateral trade examples that we discuss in the next section. It is straightforward to generalize the definitions of mechanisms and information structures so that each A_i and S_i is a measurable set, which may be uncountably infinite. Similarly, we may take strategies to be probability transition kernels, and generalize our definitions of Bayes Nash equilibrium, guarantees, and potentials in the obvious way. This is the approach adopted in Brooks and Du (2021b), and which we informally describe in the examples, where we will construct exact saddle points for the informationally-robust mechanism design problem. As we will see, not only does this allow us to exactly attain the maximum guarantee and minimum potential, but it also permits much cleaner analytical characterization and construction of the guarantee-maximizing mechanisms and potential-minimizing information structures.

3 Informationally Robust Optimal Mechanism Design

This section surveys the key results in the literature. We begin with the general theory on the structure of guarantee-maximizing mechanisms and potential-minimizing information structures. We will conclude with two examples, for which we explicitly construct saddle points: revenue guarantees in auctions and welfare guarantees in bilateral trade.

3.1 Lagrangian formulation

Following Brooks and Du (2024), to understand the relationship between the min potential and max guarantee problems, we first invoke the revelation principle two rewrite them as bilinear saddle point problems. Specifically, to calculate the potential of a finite information structure $I = (S, \sigma)$, we first apply the revelation principle (Myerson, 1981): it is without loss to maximize over direct revelation mechanisms on I that are incentive compatible and individually rational. Thus, we maximize the following Lagrangian:

$$
\begin{aligned}
L^p(S, m, \sigma, \alpha, \beta) = &\sum_{s,\theta,\omega} w(\omega, \theta) m(\omega|s) \sigma(s, \theta) \\
&+ \sum_{i, s_i'} \sum_{s,\theta,\omega} \alpha_i(s_i'|s_i) u_i(\omega, \theta)(m(\omega|s) - m(\omega|s_i', s_{-i}))\sigma(s, \theta) \\
&+ \sum_{i} \sum_{s,\theta,\omega} \beta_i(s_i) u_i(\omega, \theta) m(\omega|s) \sigma(s, \theta).
\end{aligned} \tag{3}
$$

The first line above contains the objective that the designer wants to maximize over direct revelation mechanisms $m : S \to \Delta(\Omega)$; the second line contains the incentive compatible (IC) constraints (with multipliers $\alpha_i(s_i'|s_i) \geqslant 0$) where agent i with signal s_i does not want to misreport s_i'; and the third line contains the individual rationality (IR) constraints (with multipliers $\beta_i(s_i) \geqslant 0$) where given signal s_i agent i's expected utility from truthful reporting is non-negative. By the Lagrange multiplier theorem, we have

$$
P(S, \sigma) = \min_{\alpha \geqslant 0, \beta \geqslant 0} \max_{m: S \to \Delta(\Omega)} L^p(S, m, \sigma, \alpha, \beta), \tag{4}
$$

where the minimization over α and β tells us which IC and IR constraints are binding for computing the potential.

Likewise, to calculate the guarantee of a finite mechanism $M = (A, m)$, we apply another revelation principle: it is without loss to minimize over Bayes correlated equilibria (BCE) on M. That is, we minimize the following Lagrangian:

$$
\begin{aligned}
L^g(A, m, \sigma, \alpha) = &\sum_{a,\theta,\omega} w(\omega, \theta) m(\omega|a) \sigma(a, \theta) \\
&- \sum_{i, a_i'} \sum_{a,\theta,\omega} \alpha_i(a_i'|a_i) u_i(\omega, \theta)(m(\omega|a)\sigma(a, \theta) - m(\omega|a_i', a_{-i}))\sigma(a, \theta).
\end{aligned} \tag{5}
$$

Analogous to the potential Lagrangian, the first line contains the objective that the designer wants to minimize (to seek a guarantee) over joint distributions σ of actions and states, where the actions are "recommendations" of a hypothetical mediator; the second line contains the obedience constraints of BCE (with multipliers $\alpha_i(a_i'|a_i) \geqslant 0$) where agent i when recommended to play a_i does not want to deviate to a_i'. We have

$$
G(A, m) = \max_{\alpha \geqslant 0} \min_{\sigma \in \Delta(A \times \Theta): \mathrm{marg}_\Theta \sigma = \mu} L^g(A, m, \sigma, \alpha), \tag{6}
$$

where now the maximization over α tells us which obedience constraints are binding for computing the guarantee.

In summary, the minimum potential is given by

$$
\inf_{I \in \mathcal{I}} P(I) = \inf_{(S,\sigma) \in \mathcal{I}} \min_{\alpha \geqslant 0, \beta \geqslant 0} \max_{m: S \to \Delta(\Omega)} L^p(S, m, \sigma, \alpha, \beta), \tag{7}
$$

and the maximum guarantee is given by

$$
\sup_{M \in \mathcal{M}^0} G(M) = \sup_{(A,m) \in \mathcal{M}^0} \max_{\alpha \geqslant 0} \min_{\sigma \in \Delta(A \times \Theta): \mathrm{marg}_\Theta \sigma = \mu} L^g(A, m, \sigma, \alpha). \tag{8}
$$

The optimization problems in (7) and (8) are non-linear (since σ, m and α are multiplied together in L^p and L^g) and therefore seemingly intractable to solve and obtain insights. Moreover, while the two Lagrangians seem to contain similar terms for the designer's objective and the truthtelling/obedience constraints, the latter appear with opposite signs (a plus sign in front of α_i in L^p but a minus sign in front of α_i in L^g).

3.2 Dual reductions

An important insight into the structure of the solutions to (7) and (8) comes from Brooks and Du (2025), who show that it is without loss to restrict attention to specific patterns of binding constraints and values for the Lagrange multipliers in the two problems. Specifically, they develop a "dual reduction" procedure inspired by Myerson (1997) that simultaneously simplifies mechanism/information structure and the multipliers.

To illustrate how dual reduction works, let us focus on the guarantee maximization problem. Fix a participation-secure mechanism $M = (A, m)$ and an arbitrary multiplier $\alpha_i(a_i'|a_i)$ for L^g. Moreover, since L^g is not affected by the value of $\alpha_i(a_i|a_i)$, we can without loss of generality assume that there is a large $C > 0$ such that

$$\sum_{a_i'} \alpha_i(a_i'|a_i) = C > 0$$

for every $a_i \in A_i$, .

The dual reduction procedure produces a new (reduced) mechanism $\widehat{M}$ and a new set of multipliers $\widehat{\alpha}$ for which

$$\min_{\widehat{\sigma} \in \Delta(X \times \Theta): \operatorname{marg}_\Theta \sigma = \mu} L^g(X, \widehat{m}, \widehat{\sigma}, \widehat{\alpha}(C)) \geqslant \min_{\sigma \in \Delta(A \times \Theta): \operatorname{marg}_\Theta \sigma = \mu} L^g(A, m, \sigma, \alpha). \tag{9}$$

Here is how the construction works. The actions in the reduced mechanism are non-negative integers, where each integer $l \geqslant 0$ is associated with a particular *mixed strategy* in the original mechanism M, which we denote by $b_i(\cdot|l)$. In particular, when agent i plays the action $l = 0$, this results in the mixture that puts probability one on the action $0 \in A_i$ (which we recall was a secure action in the mechanism M), that is, $b_i(0|0) = 1$. The rest of the mixtures are defined recursively according to

$$b_i(a_i|l+1) = \sum_{a_i'} \frac{\alpha_i(a_i|a_i')}{C} b_i(a_i'|l).$$

Now define the reduced mechanism $\widehat{M} = (X, \widehat{m})$ from M so that only these mixed actions are available to play: the action space for each agent is the non-negative integers

$$X_i = \{0, 1, 2, \ldots\},$$

and

$$\widehat{m}(\omega|x) = \sum_a m(\omega|a) \prod_i b_i(a_i|x_i)$$

for all $\omega \in \Omega$ and $x \in X$. Thus, in the reduced mechanism, agents choose their mixtures, actions are drawn independently, and then the original mechanism is run.

Finally, define the new multipliers $\widehat{\alpha}$ according to

$$\widehat{\alpha}_i(C)(x_i'|x_i) = \begin{cases} C & x_i' = x_i + 1, \\ 0 & x_i' \neq x_i + 1, \end{cases} \tag{10}$$

Now, why does (9) hold? For any $\widehat{\sigma} \in \Delta(X \times \Theta)$, we can define $\sigma \in \Delta(A \times \Theta)$ such that $\sigma(a, \theta) = \sum_x \widehat{\sigma}(x, \theta) \prod_i b_i(a_i|x_i)$; straightforward algebra then shows that $L^g(X, \widehat{m}, \widehat{\sigma}, \widehat{\alpha}(C)) = L^g(A, m, \sigma, \alpha)$, which proves (9).

In effect, α represents a particular pattern of deviations, where $\alpha_i(a_i'|a_i)$ is proportional to a likelihood of deviating from a_i to a_i'. Every such pattern of deviations implies a lower bound on the guarantee. The dual reduction simultaneously limits what agents can do in the mechanism (by restricting them to mixtures in the original mechanism) while also preserving the ability for the agents to deviate proportional to α, so that the implied lower bound on the guarantee can only increase.

While $\widehat{M}$ is not a finite mechanism, we can restrict $\widehat{m}$ to $X(k) = \prod_i X_i(k)$, where

$$X_i(k) = \{0, 1, \ldots, k\}.$$

Denoting this restriction as $\widehat{m}^k$ and using the same multiplier as in (10), Brooks and Du (2025) show that

$$\lim_{k \to \infty} \min_{\widehat{\sigma} \in \Delta(X(k) \times \Theta) : \mathrm{marg}_\Theta \, \sigma = \mu} L^g(X(k), \widehat{m}^k, \widehat{\sigma}, \widehat{\alpha}) = \min_{\widehat{\sigma} \in \Delta(X \times \Theta) : \mathrm{marg}_\Theta \, \sigma = \mu} L^g(X, \widehat{m}, \widehat{\sigma}, \widehat{\alpha}(C)).$$

Therefore, we conclude that the maximum guarantee may be computed using multipliers of the form in (10):

$$\sup_{M \in \mathcal{M}^0} G(M) = \sup_{k, C} \sup_{(X(k), m) \in \mathcal{M}^0} \min_{\sigma \in \Delta(X(k) \times \Theta) : \mathrm{marg}_\Theta \, \sigma = \mu} L^g(X(k), m, \sigma, \widehat{\alpha}(C)). \tag{11}$$

What this result means is that in solving for guarantee-maximizing mechanisms, it is without loss to restrict attention to mechanisms in which the actions can be *ordered* (as non-negative integers), the lowest action is secure, and the binding equilibrium constraints are those associated with deviating from an action to its successor in the sequence. We may interpret the actions as representing the strength of an agent's objections to the designer's goal, with the secure action 0 representing the strongest objection. And the multipliers $\widehat{\alpha}$ show that the constraints that matter are those associated with slightly underrepresenting one's objections. Moreover, it is without loss to take the Lagrange multiplier on the binding local obedience constraints to be constant. This is what we referred to as a *regular mechanism* in the introduction.

Brooks and Du (2025) show that there is an analogous dual reduction procedure for the information structure, and to compute the minimum potential it is without loss to use signal space

$$\overline{X}_i(k) = \{0, 1, \ldots, k\} \cup \{\infty\}$$

for some k, and multipliers of the form

$$\tilde{\alpha}_i(C)(x_i'|x_i) = \begin{cases} C & x_i' = x_i - 1, \\ 0 & x_i' \neq x_i - 1, \end{cases} \quad \tilde{\beta}_i(C)(x_i) = \begin{cases} C & x_i = 0, \\ 0 & x_i \neq 0. \end{cases} \tag{12}$$

We define $\infty - 1 = \infty$, so we are ignoring the IC (and IR) constraint of the type $x_i = \infty$. Brooks and Du (2025) prove that

$$\inf_{I \in \mathcal{I}} P(I) = \inf_{k,C} \inf_{(\overline{X}(k),\sigma) \in \mathcal{I}} \max_{m:\overline{X}(k) \to \Delta(\Omega)} L^p(\overline{X}(k), m, \sigma, \tilde{\alpha}(C), \tilde{\beta}(C)). \tag{13}$$

Given an information structure I and multipliers (α, β), the construction of the dual reduction $\tilde{I}$ proceeds somewhat differently: Each multiplier is still proportional to the likelihood of a deviation, where $\alpha_i(s_i'|s_i)$ corresponds to mimicking s_i' when the true signal is s_i, and $\beta_i(s_i)$ corresponds to leaving the mechanism. Now imagine drawing (s, θ) from I, and then propagating these stochastic mimickings and departures. Either an agent leaves the mechanism in finite time, or they never leave. In the dual reduction, agents simply observe how long it took them to leave the mechanism (or if they never left). Thus, the signal in the dual reduction is a garbling of their signal in the original mechanism. Brooks and Du (2025) show that in the dual reduction, when the signal $l > 0$ mimics the signal $l - 1$ (or leaves if $l = 0$), this replicates the stochastic deviation (α, β) in the original information structure. But now fewer outcomes are implementable in the dual reduction, since the mechanism can only depend on the departure time, not the true signals. Moreover, if the reduced signal is a censored by pooling together all departure times above a threshold, as long as the threshold is sufficiently large, the effect on the potential is negligible.

Brooks and Du (2025) refer to an information structure with this pattern of binding incentive and participation constraints $(\tilde{\alpha}, \tilde{\beta})$ as *regular*. Similar to the mechanism, the signals in $\overline{X}_i(k)$ are ordered by the strength of their objection to participation; but the binding constraint is that agents might want to slightly overstate their objections, or to not participate if they have the strongest objections.

Returning to our discussion from the introduction, we claimed that potential minimization would necessarily lead to information structures that are relatively "simple," and that this simple structure would also be reflected in guarantee-maximizing mechanisms. What we were referring to was precisely these findings, that in maximizing the guarantee and minimizing the potential, it is without loss to restrict attention to mechanisms and information structures, respectively, that are regular.

3.3 The Bounding Programs

Thus, it is without loss to restrict attention to the simple pattern of equilibrium constraints according to (10) and (12). As we noted above, the constant C can always be made larger, which corresponds to increasing $\widehat{\alpha}_i(l|l)$ or $\tilde{\alpha}_i(l|l)$, and scaling up the weight on local obedience/truthtelling constraints. In the context of the dual reduction of a mechanism, this would correspond to the actions being more "similar" to each other (since the rate of switching actions in original mechanism is inversely proportional to C).

Thus, it is natural to label the actions so that as the multiplier C becomes larger, the actions are closer together. This is the approach in Brooks and Du (2024), who consider a signal/action space $S_i = A_i = X_i(k^2)$, where $X_i(k^2)$ is now normalized as

$$X_i(k^2) = \{0, 1/k, 2/k, \ldots, k\}.$$

As $k \to \infty$, $X_i(k^2)$ converges to $\mathbb{R}_+$. Since the space between consecutive signal/action in $X_i(k^2)$ is $1/k$, we consider $C = k$ in (10) and (12) so the Lagrangians L^p and L^g involve discrete derivatives of m and σ. Loosely speaking, as $k \to \infty$, these discrete derivatives approximate derivatives of the respective objects.

Substituting (10) and (12) with $C = k$ into L^p and L^g yields:

$$\begin{aligned} L^p(X(k^2), \sigma, m, k) &\equiv L^p(X(k^2), \sigma, m, \tilde{\alpha}(k), \tilde{\beta}(k)) \\ &= \sum_{x,\theta,\omega} w(\omega,\theta) m(\omega|x)\sigma(x,\theta) + \sum_i \sum_{x,\theta,\omega} u_i(\omega,\theta)\nabla_i^- m(\omega|x)\sigma(x,\theta), \end{aligned} \tag{14}$$

$$\begin{aligned} L^g(X(k^2), \sigma, m, k) &\equiv L^g(X(k^2), \sigma, m, \widehat{\alpha}(k)) \\ &= \sum_{x,\theta,\omega} w(\omega,\theta) m(\omega|x)\sigma(x,\theta) + \sum_i \sum_{x,\theta,\omega} u_i(\omega,\theta)\nabla_i^+ m(\omega|x)\sigma(x,\theta), \end{aligned} \tag{15}$$

where

$$\nabla_i^- f(x) = \begin{cases} k(f(x_i, x_{-i}) - f(x_i - 1/k, x_{-i})) & x_i > 0, \\ k f(x_i, x_{-i}) & x_i = 0, \end{cases}$$

$$\nabla_i^+ f(x) = \begin{cases} k(f(x_i + 1/k, x_{-i}) - f(x_i, x_{-i})) & x_i < k, \\ 0 & x_i = k. \end{cases}$$

Thus, the difference in directions in $\widehat{\alpha}$ and $\tilde{\alpha}$ flips the opposing signs of the incentive constraints in L^p and L^g and results in similar-looking Lagrangians, albeit with the distinction of left-hand vs. right-hand derivatives. The two Lagrangians $L^p(X(k^2), \sigma, m, k)$ and $L^g(X(k^2), \sigma, m, k)$ should be close to each other if k is large and m and σ are smooth functions of x. This statement is made precise in Proposition 6 of Brooks and Du (2024).

Since they are obtained by restricting the problems in (7) and (8), $L^p(X(k^2), \sigma, m, k)$ and $L^g(X(k^2), \sigma, m, k)$ yield upper and lower bounds on the potential and guarantee:

$$\overline{P}(X(k^2), \sigma) = \max_{m: X(k^2) \to \Delta(\Omega)} L^p(X(k^2), \sigma, m, k) \geqslant P(X(k^2), \sigma) \tag{16}$$

$$\underline{G}(X(k^2), m) = \min_{\sigma \in \Delta(X(k^2) \times \Theta): \mathrm{marg}_\Theta \sigma = \mu} L^g(X(k^2), \sigma, m, k) \leqslant G(X(k^2), m), \tag{17}$$

Optimizing these bounds yields the bounding programs:

$$\overline{P}(X(k^2)) = \min_{\sigma \in \Delta(X(k^2) \times \Theta): \mathrm{marg}_\Theta \sigma = \mu} \overline{P}(X(k^2), \sigma), \tag{18}$$

$$\underline{G}(X(k^2)) = \max_{m:X(k^2)\to\Delta(\Omega)} \underline{G}(X(k^2), m), \text{st:} \sum_{\omega} u_i(\omega,\theta)m(\omega|0,a_{-i}) \geqslant 0 \; \forall i, \theta, a_{-i}, \tag{19}$$

with

$$\underline{G}(X(k^2)) \leqslant \sup_{M\in\mathcal{M}^0} G(M) = G^* \leqslant P^* = \inf_{I\in\mathcal{I}} P(I) \leqslant \overline{P}(X(k^2)). \tag{20}$$

The result described in the previous section on dual reductions shows that as k goes to infinity, $\underline{G}(X(k^2))$ converges to G^* and $\overline{P}(X(k^2))$ converges to P^*. Moreover, solutions to the bounding program (18) are approximate potential minimizers, in that the potential of the solution converges to P^* as $k \to \infty$. Similarly, the guarantee of the solution to (19) converges to G^* as $k \to \infty$. An important remaining question, though, is whether $G^* = P^*$, that is, whether or not there is a duality gap.

Brooks and Du (2024) show, for a class of problems involving revenue maximization in multi-good auctions, that as k becomes large, the value of the bounding programs must be close to one another. This proves that for those problems, the duality gap is zero. To gain some intuition into this result, observe that $L^p(X(k^2), \sigma, m, k)$ and $L^g(X(k^2), \sigma, m, k)$ are bi-linear functions of σ and m, where they live in compact and convex spaces. Therefore, if $L^p(X(k^2), \sigma, m, k)$ and $L^g(X(k^2), \sigma, m, k)$ are close to each other when k is large, then Sion's minimax theorem would allow us to switch max and min and would imply that $\overline{P}(X(k^2))$ and $\underline{G}(X(k^2))$ must be close to each other.

Finally, we have

$$L^g(X(k^2), m, \sigma, k) = \sum_{x,\theta} \lambda(x,\theta)\sigma(x,\theta),$$

where

$$\lambda(x,\theta) = \sum_{\omega} w(\omega,\theta)m(\omega|x) + \sum_i \sum_{\omega} u_i(\omega,\theta)\nabla_i^+ m(\omega|x). \tag{21}$$

We call $\lambda(x,\theta)$ the *strategic virtual objective* (SVO) for the mechanism $(X(k^2), m)$. We have:

$$\underline{G}(X(k^2), m) = \min_{\sigma\in\Delta(X(k)\times\Theta):\text{marg}_\Theta\,\sigma=\mu} L^g(X(k^2), \sigma, m, k) = \sum_{\theta} \mu(\theta) \min_x \lambda(x,\theta). \tag{22}$$

Thus, a mechanism maximizes the guarantee if it maximizes the expected (across states) lowest (across action profiles) strategic virtual objective. In effect, the strategic virtual objective represents the designer's objective, plus extra terms capturing the agents local incentives to underrepresent their objections to the designer's goals.

Likewise, using summation by parts, we have

$$L^p(X(k^2), \sigma, m, k) = \sum_{x,\omega} \gamma(x,\omega)m(\omega|x),$$

where

$$\gamma(x,\omega) = \sum_{\theta} w(\omega,\theta)\sigma(x,\theta) - \sum_i \sum_{\theta} u_i(\omega,\theta)\tilde{\nabla}_i^+ \sigma(x,\theta), \tag{23}$$

and

$$\tilde{\nabla}_i^+ f(x) = \begin{cases} -kf(x_i, x_{-i}) & x_i = k \\ k(f(x_i + 1/k, x_{-i}) - f(x_i, x_{-i})) & x_i < k \end{cases}$$

We call $\gamma(x, \omega)$ the *informational virtual objective* (IVO) for the information structure $(X(k^2), \sigma)$. We have:

$$\overline{P}(X(k^2), \sigma) = \max_{m: X(k) \to \Delta(\Omega)} L^p(X(k^2), \sigma, m, k) = \sum_x \max_\omega \gamma(x, \omega). \tag{24}$$

An information structure minimizes the potential if it minimizes the expected (across action profiles) maximum (across outcomes) informational virtual objective. In effect, the informational virtual objective is the designer's objective, plus additional terms representing the agents' local incentives to overrepresent their objections.

While the strategic virtual objective is a relatively new concept, the informational virtual objective is in fact a generalization of the virtual value introduced in Myerson (1981). We return to this in the auctions example.

3.4 Saddle Points

Heretofore, we have maintained finiteness of the mechanisms and information structures. But as we transition to examples, it will be more convenient to work with the continuum analogue of the programs $\overline{P}(X(k^2), \sigma)$ and $\underline{G}(X(k^2), m)$, so that the max guarantee and min potential may be attained exactly. In the limit, $X(k^2)$ "fills in" the non-negative real line, and the the discrete derivatives in the bounding programs would become directional derivatives. By working with the differential form of the bounding program, we can avail ourselves of the calculus to solve the programs.

All of this sounds quite speculative and lacking in rigor. However, there is an entirely rigorous way to approach this issue. Suppose we have a candidate saddle point (M, I), where the action space and signal space are $\mathbb{R}_+$ (but still taking Θ and Ω to be finite). Suppose further that the candidate $M = (\mathbb{R}_+^N, m)$ is such that $\nabla_i m(\omega|x)$ exists for all i and x, where $\nabla_i = \partial/\partial x_i$. Then, a necessary condition for equilibrium is that local upward obedience constraints are satisfied. Weak duality then implies a lower bound on the guarantee of the mechanism M:

$$G(M) \geqslant \underline{G}(M) \equiv \sum_\theta \mu(\theta) \inf_x \lambda(x, \theta)$$

where (cf. equation (21))

$$\lambda(x, \theta) = \sum_\omega w(\omega, \theta) m(\omega|x) + \sum_i \sum_\omega u_i(\omega, \theta) \nabla_i m(\omega|x).$$

Similarly, suppose I is such that the joint distribution over (x, θ) is absolutely continuous in x, so that we may write it as $\sigma(x, \theta)dx$. Suppose further that $\sigma(x, \theta)$ is differentiable in

x_i for all θ. Then, as local downward truthtelling constraints are a necessary condition for equilibrium, weak duality implies that

$$P(I) \leqslant \overline{P}(I) \equiv \int_x \max_\omega \gamma(x,\omega)dx,$$

where (cf. equation (23))

$$\gamma(x,\omega) = \sum_\theta w(\omega,\theta)\sigma(x,\theta) - \sum_i \sum_\theta u_i(\omega,\theta)\nabla_i \sigma(x,\theta).$$

Thus, one way to certify that (M, I) is in fact a saddle point would be to show that

$$\sum_\theta \mu(\theta) \inf_x \lambda(x,\theta) = \int_x \max_\omega \gamma(x,\omega)dx.$$

This is precisely the approach adopted in Brooks and Du (2021b,a, 2023) and Brooks, Du, and Feffer (2025).

Notice that what we have just described is a "guess and verify" approach to solving the informationally-robust design problem. This is to be contrasted with the more systematic approach we previously described, via the bounding linear programs, whose solutions necessarily converge to the respective optimal values, whether or not there is a duality gap and whether or not there are saddle points. The guess and verify approach, by contrast, can succeed only if a saddle point exists and there is no duality gap.

Nonetheless, for problems in auctions, public goods, bilateral trade, and market exchange, the guess and verify approach has yielded results. The key step is to generate informed guesses for the saddle point (M, I), which may be verified via weak duality. One way to do so is to utilize the bounding programs (18) and (19), which are are finite dimensional linear programs (cf. equations (22) and (24)) and hence relatively computationally tractable. As in Roger Myerson's anecdote from the beginning of this article, numerical solutions of (18) and (19) can be used to generate and test conjectures for the guarantee-maximizing mechanism and potential-minimizing information in applications, and infer functional forms in the continuum limit.

Another way to guess saddle points is using linear programming duality. In particular, as argued in Brooks and Du (2024), the bounding programs (18) and (19) are an "approximate" dual pair of linear programs, except that the direction of equilibrium constraints and the form of the participation constraint are different. In this dual pairing, the likelihood $m(\omega|x)$ plays the role of a Lagrange multiplier on the constraint that $\gamma(x,\omega) \leqslant \max_{\omega'} \gamma(x,\omega')$. And $\sigma(x,\theta)$ plays the role of Lagrange multiplier on the constraint that $\lambda(x,\theta) \geqslant \min_{x'} \lambda(x',\theta)$. A natural guess is that in the continuum limit, this approximate duality is exact, so that a saddle point would satisfy *complementary slackness*[7]:

$$m(\omega|x) > 0 \Longrightarrow \omega \in \operatorname{argmax}_{\omega'} \gamma(x,\omega'), \tag{25}$$

[7]Here is another way to think about complementary slackness. Suppose $L(\sigma,m) = \int_x \sum_\omega \gamma(x,\omega)m(\omega|x)dx = \sum_\theta \int_x \lambda(x,\theta)\sigma(x,\theta)dx$. Consider a zero-sum game where player 1 chooses σ (such that $\operatorname{marg}_\Theta \sigma = \mu$) to minimize $L(\sigma,m)$, while player 2 chooses m to maximize $L(\sigma,m)$ (cf. the minmax and maxmin problems in (18) and (19)). Then the complementary slackness (25) and (26) are exactly the conditions for (σ,m) to be a Nash equilibrium of this game.

$$\sigma(x, \theta) > 0 \Longrightarrow x \in \operatorname{argmin}_{x'} \lambda(x', \theta). \tag{26}$$

There is no formal result in this literature that establishes complementary slackness as a necessary condition for a saddle point. However, these complementary slackness conditions turn out to hold in the saddle points that have been identified in the literature. Moreover, the ansatz of complementary slackness has proven extremely valuable in solving for saddle points in applications, as we illustrate in the next two sections.

3.5 Example: A Common Value Auction

We now describe a simple version of the common value auction problem of Brooks and Du (2021b). Suppose there is a single unit of common-value good for sale. For simplicity, suppose the common value is either 0 or 1, i.e., $\Theta = \{0, 1\}$, with distribution $\mu \in \Delta(\Theta)$. The outcome space is $\Delta(\Omega)$ where $\Omega = \{0, 1, \ldots, N\} \times \{-t_{max}, t_{max}\}^N$. For $\omega = (\iota, \tau_1, \ldots, \tau_N) \in \Omega$ and $\theta \in \Theta$,

$$u_i(\theta, \omega) = \theta \mathbb{I}_{i=\iota} - \tau_i,$$

and

$$w(\theta, \omega) = \sum_i \tau_i.$$

Thus $\iota = 0$ means the good is not sold; $\iota \in \{1, 2, \ldots, N\}$ means the good is sold to agent ι. And τ_i is the monetary transfer from agent i to the designer. The designer's preference is total transfer, i.e., revenue.

Note that the transfer τ_i can only take on two values. But randomizing over these values induces an expected net payment, which we interpret as the continuous transfer typically modeled in auction theory. We will consider the case when t_{max} is large, so that transfer is essentially unbounded.[8] Specifically, for an $\tilde{\omega} \in \Delta(\Omega)$, we denote the allocation probability $q_i = \tilde{\omega}(\{(\iota, \tau_1, \ldots, \tau_N) \in \Omega : \iota = i\})$, and the (expected) transfer $t_i = (2\tilde{\omega}(\{(\iota, \tau_1, \ldots, \tau_N) \in \Omega : \tau_i = t_{max}\}) - 1)t_{max}$. Then we have

$$\sum_\omega u_i(\theta, \omega)\tilde{\omega}(\omega) = \theta q_i - t_i,$$

and

$$\sum_\omega w(\theta, \omega)\tilde{\omega}(\omega) = \sum_i t_i.$$

So, we may reparametrize the outcome, and instead of choosing $\tilde{\omega} \in \Delta(\Omega)$, we will model the outcome in reduced form as $q \in \Delta(\{0, 1, \ldots, N\})$ and $t \in [-t_{max}, t_{max}]^N$. Note that when we sum over i we always meant for $i \in \{1, 2, \ldots, N\}$, so $q_0 = 1 - \sum_i q_i$.

[8] Brooks and Du (2021b) formally consider the case where the transfer is unbounded. This is not formally subsumed within our model with finitely many outcomes. Brooks and Du (2024) derive analogues of the bounding programs when transfers are free variables. However, for the present purposes, it suffices to consider the case where transfers can be large but finite.

3.5.1 Potential-minimizing information structure

The IVO of an information structure $(\mathbb{R}_+^N, \sigma)$ is

$$\gamma(x,(q,t)) \equiv \sum_{\omega} \gamma(x,\omega)\tilde{\omega}(\omega) = \sum_i \sum_\theta \left(t_i \sigma(x,\theta) - \nabla_i \sigma(x,\theta)(\theta q_i - t_i)\right).$$

For every x, we maximize $\gamma(x,(q,t))$ over (q,t), clearly we want to set t_i to be t_{max} (respectively, $-t_{max}$) if the coefficient of t_i in $\gamma(x,(q,t))$ is positive (respectively, negative). If transfers were unbounded, the coefficient of t_i would have to be zero:

$$\sum_\theta (\sigma(x,\theta) + \nabla_i \sigma(x,\theta)) = 0 \tag{27}$$

for every i and $x \in \mathbb{R}_+^N$. In the version of the model we are describing, the transfer is bounded, but the bounds are as large as we would like them to be. So, it is natural to consider information structures that satisfy (27). We shall subsequently verify that when t_{max} is sufficiently large, such information structures do in fact minimize the potential.

Denote the marginal distribution of σ on x as $\rho(x) = \sum_\theta \sigma(x,\theta)$. Then (27) implies that for every $x_i \in \mathbb{R}_+$ and $x_{-i} \in \mathbb{R}_+^{N-1}$

$$\rho(x) + \nabla_i \rho(x) = 0,$$

so that $\rho(x) = \rho(0, x_{-i})e^{-x_i}$. Iterating this across agents, and using the condition that $\int_{x_i \in \mathbb{R}_+} \rho(x_i|x_{-i})dx_i = 1$ we conclude that implies that the the agents' signals must be independently and exponentially distributed! For future references we denote this marginal distribution as $\overline{\rho}(x) \equiv e^{-\Sigma x}$, where $\Sigma x \equiv \sum_i x_i$.

We now consider σ with $\overline{\rho}$ as the marginal on x and substitute it into the IVO. Denote the interim expected value as

$$v(x) = \frac{\sum_\theta \theta \sigma(x,\theta)}{\overline{\rho}(x)}. \tag{28}$$

Then the IVO can be written as

$$\begin{aligned} \gamma(x,(q,t)) &= -\sum_i \sum_\theta \nabla_i \sigma(x,\theta)\theta q_i \\ &= -\sum_i \nabla_i(\overline{\rho}(x)v(x))q_i \\ &= \overline{\rho}(x)\sum_i (v(x) - \nabla_i v(x))q_i, \end{aligned} \tag{29}$$

where we used the fact that $\nabla_i \overline{\rho}(x) = -\overline{\rho}(x)$. Since the inverse hazard rate of a standard exponential distribution is one, this is exactly the Myersonian formula for the virtual value, as generalized to interdependent values by Bulow and Klemperer (1996). Clearly, to maximize $\gamma(x,(q,t))$ over q, we should have $q_i > 0$ only if i's virtual value $v(x) - \nabla_i v(x)$ is non-negative and maximal among all agents. This immediately suggests that to minimize $\overline{P}(\mathbb{R}_+^N, \sigma)$, we should have $v(x) - \nabla_i v(x) = 0$ for all i for a large set of x to keep the maximum IVO as small as possible, i.e., $v(x) = Ae^{\Sigma x}$ for some constant $A > 0$. Since we must have $v(x) \leqslant 1$, this suggests an interim value function $\overline{v}(x) = \min\{Ae^{\Sigma x}, 1\}$, so that

$$\gamma(x,(q,t)) = \begin{cases} 0 & \Sigma x \leqslant \overline{y}, \\ \overline{\rho}(x)(1-q_0) & \Sigma x > \overline{y}, \end{cases} \tag{30}$$

where $\overline{y} = \log(1/A)$.

The constant A is chosen so that the marginal distribution over Θ is μ, i.e.,

$$\int_{x \in \mathbb{R}_+^N} \min\{Ae^{\Sigma x}, 1\}\overline{\rho}(x)dx = \mu(1). \tag{31}$$

The sum of the signals has an Erlang distribution, with density function

$$g_N(y) = \frac{y^{N-1}e^{-y}}{(N-1)!} \tag{32}$$

and cumulative distribution

$$G_N(y) = \int_{z=0}^{y} g_N(z)dz = 1 - \sum_{l=1}^{N} g_l(y). \tag{33}$$

Thus equation (31) can be rewritten as

$$\begin{aligned} \int_{y=0}^{\overline{y}} Ae^{y} g_N(y) + 1 - G_N(\overline{y}) &= A\frac{\overline{y}^N}{N!} + 1 - G_N(\overline{y}) \\ &= g_{N+1}(\overline{y}) + 1 - G_N(\overline{y}) \\ &= 1 - G_{N+1}(\overline{y}) = \mu(1), \end{aligned}$$

i.e.,

$$G_{N+1}(\overline{y}) = \mu(0). \tag{34}$$

Summing up, the candidate for the potential-minimizing information structure is $\overline{I} = (\mathbb{R}_+^N, \overline{\sigma})$, where

$$\overline{\sigma}(x,1) = \overline{\rho}(x)\overline{v}(x), \quad \overline{\sigma}(x,0) = \overline{\rho}(x)(1-\overline{v}(x)),$$

which has a potential of at most

$$\int_{x \in \mathbb{R}_+^N} \max_{q,t} \gamma(x,(q,t))\, dx = 1 - G_N(\overline{y}). \tag{35}$$

3.5.2 Guarantee-maximizing mechanism

Consider a mechanism $(\mathbb{R}^N_+, q, t)$ where $q : \mathbb{R}^N_+ \to \Delta(\{0, 1, \dots, N\})$ and $t : \mathbb{R}^N_+ \to [-t_{max}, t_{max}]^N$. Its SVO is:

$$\lambda(\theta, x) = \sum_i (t_i(x) + \theta \nabla_i q_i(x) - \nabla_i t_i(x)).$$

Since both $\overline{\sigma}(x, 0) > 0$ and $\overline{\sigma}(x, 1) > 0$ when $\Sigma x \leqslant \overline{y}$, but $\overline{\sigma}(x, 0) = 0$ and $\overline{\sigma}(x, 1) > 0$ when $\Sigma x > \overline{y}$, by complementary slackness condition (26), we look for a mechanism for which all action profiles minimize the SVO $\theta = 1$, and action profiles with $\Sigma x \leqslant \overline{y}$ minimize the SVO when $\theta = 0$. Equivalently, $\lambda(1, x) = L_1$ for all x, $\lambda(0, x) = L_0$ when $\Sigma x \leqslant \overline{y}$, and $L(0, x) \geqslant L_0$ when $\Sigma x > \overline{y}$, for constants L_1 and L_0. This implies that

$$\nabla \cdot q(x) \equiv \sum_i \nabla_i q_i(x) \begin{cases} = B & \Sigma x \leqslant \overline{y}, \\ \leqslant B & \Sigma x > \overline{y}, \end{cases}$$

for some constant $B > 0$, and

$$L_1 = L_0 + B.$$

Moreover, when $\Sigma x > \overline{y}$, the IVO of $\overline{\sigma}$ is not maximized when $q_0 > 0$ (see equation (30)), so by complementary slackness condition (25) we must have

$$\sum_i q_i(x) = 1, \quad \Sigma x > \overline{y}.$$

A solution to this PDE is

$$\overline{q}_i(x) = \frac{x_i}{\max(\Sigma x, \overline{y})},$$

which gives

$$\nabla \cdot \overline{q}(x) = \nabla \cdot \overline{q}(\Sigma x) = \begin{cases} N/\overline{y} = B & \Sigma x \leqslant \overline{y}, \\ (N-1)/\Sigma x & \Sigma x > \overline{y}. \end{cases}$$

In fact, in the case of $N = 2$, if we want $\nabla \cdot q(x)$ to be a function of Σx, then one can show that the above $\overline{q}$ is the only possibility.

Now given a proportional allocation, it is natural to conjecture that the transfer function should be proportional as well:

$$\overline{t}_i(x) = \frac{x_i}{\Sigma x} T(\Sigma x)$$

for some total transfer function $T : \mathbb{R}_+ \to \mathbb{R}$. Brooks and Du (2021b) call a mechanism $\overline{M} = (\mathbb{R}^N_+, \overline{q}, \overline{t})$ of this form a *proportional auction*.

Then we have

$$\sum_i (\overline{t}_i(x) - \nabla_i \overline{t}_i(x)) = T(\Sigma x) - \frac{N-1}{\Sigma x} T(\Sigma x) - T'(\Sigma x)$$

$$= \lambda(1,x) - \nabla \cdot \overline{q}(\Sigma x) = L_1 - \nabla \cdot \overline{q}(\Sigma x),$$

The initial condition is $T(0) = 0$, which ensures participation security. Solving this differential equation gives

$$T(y) = \frac{\int_{z=0}^{y} (\nabla \cdot \overline{q}(z) - L_1) g_N(z) dz}{g_N(y)},$$

where g_N defined in (32).

In summary, the guarantee of proportional auction satisfies

$$G(\overline{M}) \geqslant \sum_{\theta} \mu(\theta) \min_{x} \lambda(\theta, x) = \mu(1)L_1 + \mu(0)L_0.$$

Clearly we want to make L_1 and L_0 as large as possible. However, if $L_1 > \int_{z=0}^{\infty} \nabla \cdot \overline{q}(z) g_N(z) dz$, then $\lim_{y\to\infty} T(y) = -\infty$ since $\lim_{y\to\infty} g_N(y) = 0$, which preclude the existence of any equilibrium (the agents would race to bid an ever larger x_i). Thus, we set

$$\begin{aligned} L_1 = \int_{y=0}^{\infty} \nabla \cdot \overline{q}(y) g_N(y) dy &= G_N(\overline{y}) \frac{N}{\overline{y}} + \int_{y=\overline{y}}^{\infty} \frac{N-1}{y} g_N(y) dy \\ &= G_N(\overline{y}) \frac{N}{\overline{y}} + 1 - G_{N-1}(\overline{y}), \end{aligned} \tag{36}$$

since $g_{N-1}(y) = g_N(y)\frac{N-1}{y}$.

Since $L_0 = L_1 - \frac{N}{\overline{y}}$, the proportional auction has a guarantee of

$$\begin{aligned} G(\overline{M}) \geqslant \mu(1)L_1 + \mu(0)L_0 &= G_N(\overline{y}) \frac{N}{\overline{y}} + 1 - G_{N-1}(\overline{y}) - \mu(0)\frac{N}{\overline{y}} \\ &= g_{N+1}(\overline{y}) \frac{N}{\overline{y}} + 1 - G_{N-1}(\overline{y}) \\ &= 1 - G_N(\overline{y}) \end{aligned} \tag{37}$$

where we have used equations (33) and (34). The lower bound on the guarantee in (37) is exactly equal to the upper bound on the potential in (35). Thus, $(\overline{M}, \overline{I})$ is a saddle point.

Brooks and Du (2021b) show the striking fact that as $N \to \infty$, the guarantee/potential in (37) converges to the expected common value $\mu(1)$ at the convergence rate of $O(1/\sqrt{N})$; this result also holds for any prior distribution μ of common values and generalizes an earlier result of Du (2018). Thus, the guarantee of the proportional auction converges to the first best full surplus as the market gets large, regardless of how information changes as we add more agents to the market. Such an asymptotic full surplus extraction cannot be obtained by a standard auction like the first price auction, as shown by Bergemann, Brooks, and Morris (2017, 2019).

3.6 Example: bilateral trading

Our second example is a derived from the public goods problem in Brooks and Du (2023) (see also Brooks and Du, 2024, Section 4)).

There are two agents: agent 1 is the seller with a unit of a good; agent 2 is the buyer. The state is either high or low: In the low state $\theta = (0, g)$, so the seller's value for the good is $\theta_1 = 0$, while the buyer's value is $\theta_2 = g > 0$. In the high state, $\theta = (h, h + g)$, so the seller's value for the good is $\theta_1 = h > 0$, while the buyer's value is $\theta_2 = h + g$. Thus, there is common knowledge that the gains from trade are equal to $g > 0$. Both states are equally likely, i.e., $\mu(0, g) = \mu(h, h + g) = 1/2$.

The outcome space is $\Delta(\Omega)$, where $\Omega = \{0, 1\} \times \{-t_{max}, t_{max}\}$, which represents whether or not trade takes place and the net transfer from the buyer to the seller. As in the common value auction example we look at the case where t_{max} is large. Preferences again have the quasilinear form: For $\omega = (\iota, \tau) \in \Omega$,

$$\begin{aligned} u_1(\theta, \omega) &= \tau - \theta_1 \mathbb{I}_{\iota=1}, \\ u_2(\theta, \omega) &= -\tau + \theta_2 \mathbb{I}_{\iota=1}, \\ w(\theta, \omega) &= \mathbb{I}_{\iota=1}(\theta_2 - \theta_1) = \mathbb{I}_{\iota=1} g. \end{aligned}$$

Thus, the designer's objective is to maximize gains from trade.

For an $\tilde{\omega} \in \Delta(\Omega)$, we denote the trading probability $q = \tilde{\omega}(\{(\iota, \tau) \in \Omega : \iota = 1\})$, and the (expected) transfer $t = (2\tilde{\omega}(\{(\iota, \tau) \in \Omega : \tau = t_{max}\}) - 1)t_{max}$. Then we have

$$\begin{aligned} \sum_\omega u_1(\theta, \omega)\tilde{\omega}(\omega) &= t - \theta_1 q, \\ \sum_\omega u_2(\theta, \omega)\tilde{\omega}(\omega) &= -t + \theta_2 q, \\ \sum_\omega w(\theta, \omega)\tilde{\omega}(\omega) &= q(\theta_2 - \theta_1) = qg. \end{aligned}$$

So instead of $\tilde{\omega} \in \Delta(\Omega)$ we will work directly with $q \in [0, 1]$ and $t \in [-t_{max}, t_{max}]$.

We assume that $h > g$ to rule out a trivial case in which it is always possible to implement efficient trade: If $h \leqslant g$, then any price lower than g and higher than h would be acceptable to both the buyer and the seller regardless of their information about the state.

3.6.1 Potential-minimizing information structure

The IVO of an information structure $(\mathbb{R}^N_+, \sigma)$ is

$$\gamma(x, (q, t)) \equiv \sum_\omega \gamma(x, \omega)\tilde{\omega}(\omega) = \sum_\theta \left(gq\sigma(x, \theta) - \nabla_1 \sigma(x, \theta)(t - q\theta_1) - \nabla_2 \sigma(x, \theta)(-t + q\theta_2)\right).$$

As with the auction, we may conjecture that when t_{max} is large, the coefficient of t in $\gamma(x, (q, t))$ must be zero. This is equivalent to the PDE

$$\sum_\theta (-\nabla_1 \sigma(x, \theta) + \nabla_2 \sigma(x, \theta)) = 0,$$

i.e., the marginal over x is a function of $x_1 + x_2$,

$$\sum_{\theta} \sigma(x, \theta) \equiv \rho(x) = \rho(x_1 + x_2). \tag{38}$$

Now let $\nu(x)$ be the interim probability of the high state:

$$\nu(x) = \frac{\sigma(x, (h, h+g))}{\rho(x_1 + x_2)}.$$

Substituting this back to IVO, we have

$$\begin{aligned}\gamma(x, (q, t)) &= \big(g\rho(x_1 + x_2) + \nabla_1(\rho(x_1 + x_2)\nu(x))h - \nabla_2(\rho(x_1 + x_2)\nu(x))(h + g) \\ &\quad - \nabla_2(\rho(x_1 + x_2)(1 - \nu(x)))g\big)q \\ &\quad - \nabla_2\rho(x_1 + x_2)g\big)q \\ &= \big(g\rho(x_1 + x_2) + \rho(x_1 + x_2)(\nabla_1\nu(x) - \nabla_2\nu(x))h - \rho'(x_1 + x_2)g\big)q.\end{aligned}$$

Numerical solution suggests that the minimizing $\nu(x)$ is

$$\overline{\nu}(x) = \frac{x_2}{x_1 + x_2}. \tag{39}$$

That is, conditional on the signal sum $x_1 + x_2$, a higher x_1 for the seller is a bad news about the value, while a higher x_2 for the buyer is a good news. With this ν, the IVO is a function of just $x_1 + x_2$:

$$\gamma(x, (q, t)) = \left(g(\rho(x_1 + x_2) - \rho'(x_1 + x_2)) - \frac{\rho(x_1 + x_2)}{x_1 + x_2}h\right) q.$$

As in the common value auction example, to make $\max_{q,t} \gamma(x, (q, t))$ as small as possible, we should pick ρ so that the IVO is zero:

$$g(\rho(x_1 + x_2) - \rho'(x_1 + x_2)) - \frac{\rho(x_1 + x_2)}{x_1 + x_2}h = 0. \tag{40}$$

Solving this differential equation gives $\rho(y) = Ae^y y^{-h/g}$ for some constant $A > 0$. However, unlike the information structure in the common value auction example, we have $\lim_{y\to\infty} \rho(y) \neq 0$ so we must truncate this density:

$$\overline{\rho}(y) = \begin{cases} Ae^y y^{-h/g} & y \leqslant \overline{y}, \\ 0 & y > \overline{y}, \end{cases}$$

where $\overline{y}$ will be determined later (equation (43)), and A is a normalizing constant.

For various technical reasons (one of them is that if $h/g \geqslant 2$, $\overline{\rho}(y)$ is not integrable around $y = 0$), we avoid the signal profile $(0, 0)$ and work with $\overline{I}_\epsilon = ([\epsilon/2, \infty)^2, \overline{\sigma})$, where

$$\overline{\sigma}(x, (h, h+g)) = \overline{\rho}(x_1 + x_2)\frac{x_2}{x_1 + x_2},\ \overline{\sigma}(x, (0, g)) = \overline{\rho}(x_1 + x_2)\frac{x_1}{x_1 + x_2},$$

and

$$A = \frac{1}{\int_{(x_1,x_2)\in[\epsilon/2,\infty)^2} e^{x_1+x_2}(x_1+x_2)^{-h/g}dx_1dx_2} = \frac{1}{\int_{y=\epsilon}^{\overline{y}}(y-\epsilon)e^y y^{-h/g}dy}.$$

It would seem that the potential of $\overline{I}_\epsilon$ is zero, since by construction $\gamma(x,(q,t)) = 0$. While we do have $\gamma(x,(q,t)) = 0$ for all q when $x_1 + x_2 < \overline{y}$, the density $\overline{\rho}$ is discontinuous at $x_1 + x_2 = \overline{y}$, so the derivative $\nabla_i\overline{\rho}(x_1 + x_2)$ in $\gamma(x,(q,t))$ is not well defined when $x_1 + x_2 = \overline{y}$. To work out $\gamma(x,(q,t))$ when $x_1 + x_2 = \overline{y}$, let us go back to the discrete approximation: let

$$\sigma_k(x,\theta) = B_k\overline{\sigma}(x,\theta)/k^2$$

for $x \in (X_i(k^2) \cap [\epsilon/2,\infty))^2$, where B_k is a normalization constant so that $\sum_{x,\theta}\sigma_k(x,\theta) = 1$ when summed over $x \in (X_i(k^2) \cap [\epsilon/2,\infty))^2$ and $\theta \in \Theta$. Then using the discrete IVO in (23) with $\sigma = \sigma_k$, if $x_1 + x_2 \leqslant \overline{y}$ but $x_1 + x_2 + 1/k > \overline{y}$, then

$$\gamma(x,(q,t)) = \sum_\theta \left(gqB_k\overline{\sigma}(x,\theta)/k^2 - k(0 - B_k\overline{\sigma}(x,\theta)/k^2)gq\right),$$

which is clearly maximized when $q = 1$. As $k \to \infty$, we have $B_k \to 1$, and there are approximately $k(\overline{y} - \epsilon)$ such x, so $\sum_x \max_{q,t}\gamma(x,(q,t))$ summed over these x's tends to

$$(\overline{y}-\epsilon)\overline{\rho}(\overline{y})g = \frac{(\overline{y}-\epsilon)e^{\overline{y}}\overline{y}^{-h/g}g}{\int_{y=\epsilon}^{\overline{y}}(y-\epsilon)e^y y^{-h/g}dy}.$$

Working directly with $\overline{I}_\epsilon$ but using essentially the same argument as above, Brooks and Du (2023) show that

$$P(I_\epsilon) \leqslant \frac{(\overline{y}-\epsilon)e^{\overline{y}}\overline{y}^{-h/g}g}{\int_{y=\epsilon}^{\overline{y}}(y-\epsilon)e^y y^{-h/g}dy}.$$

Therefore

$$\lim_{\epsilon\to 0} P(I_\epsilon) \leqslant \begin{cases} \frac{e^{\overline{y}}\overline{y}^{1-h/g}g}{\int_{y=0}^{\overline{y}} e^y y^{1-h/g}dy} & h/g < 2, \\ 0 & h/g \geqslant 2. \end{cases} \tag{41}$$

Thus, if $h/g > 2$, there is practically no trade under information structure I_ϵ in any budget-balanced mechanism and any equilibrium, even though there is a common knowledge of gains from trade.

In comparison, consider Akerlof (1970)'s lemons information structure where the seller knows the state and the buyer has no information. The condition $h/g > 2$ is equivalent to the Akerlof (1970) condition $(h+g)/2 + g/2 < h$ (the expected buyer value less than the seller's high value) for the breakdown of trade in the high state in the lemon information structure; nonetheless, efficient trade would still take place in the low state, i.e., with 1/2 ex-ante probability. In a sense, we have learned that efficient trade can be much harder to achieve than suggested by the lemons model.

If $h/g < 2$, then the upper bound on potential in (41) is positive. We can determine $\overline{y}$ by minimizing

$$\frac{e^{\overline{y}}\overline{y}^{1-h/g}g}{\int_{y=0}^{\overline{y}} e^y y^{1-h/g} dy},$$

though this can be done indirectly by characterizing the guarantee-maximizing mechanism (see equation (43)).

3.6.2 Guarantee-maximizing mechanism

When $h/g \geqslant 2$, the minimum potential is zero. Clearly, any mechanism has a guarantee of at least 0, so trivially obtains the maximum guarantee.

Suppose $h/g < 2$. For a mechanism $(\mathbb{R}_+^2, q, t)$, where $q : \mathbb{R}_+^2 \to [0,1]$ and $t : \mathbb{R}_+^2 \to [-t_{max}, t_{max}]$, the SVO is:

$$\lambda(\theta, x) = q(x)g + \nabla_1 t(x) - \theta_1 \nabla_1 q(x) - \nabla_2 t(x) + \theta_2 \nabla_2 q(x).$$

Since $\overline{\sigma}(x, \theta) > 0$ for both high and low θ for all x such that $x_1 + x_2 \leqslant \overline{y}$, by complementary slackness condition (26), we want $\lambda(\theta, x) = L_\theta$ when $x_1 + x_2 \leqslant \overline{y}$, and $\lambda(\theta, x) \geqslant L_\theta$ when $x_1 + x_2 > \overline{y}$, where L_θ's are constants. Moreover, since $\gamma(x, (q, t))$ is uniquely maximized by $q = 1$ when $x_1 + x_2 = \overline{y}$, by complementary slackness condition (25) we want $q(x) = 1$ when $x_1 + x_2 = \overline{y}$.

Since $x_1 + x_2$ is a "sufficient statistic" for the signal profile in $\overline{I}_\epsilon$, we conjecture that $q(x) = q(x_1 + x_2)$. This conjecture is confirmed by the numerical solutions of program (19), which also suggest

$$t(x) = q(x_1 + x_2)\left(g + (h - g)\frac{x_2}{x_1 + x_2}\right).$$

The above transfer rule is participation secure: with $x_1 = 0$ the seller can guarantee to sell at his high price of h, and with $x_2 = 0$ the buyer can guarantee to buy at his low price of g. (We will require $q(0) = 0$ so it is irrelevant how the price is defined when $x = (0, 0)$.) In general the pricing function interpolates between g and h, the higher x_i is, the more favorable the price is to the other agent $j \neq i$. That is the cost of raising x_i; the benefit is a higher trading probability, since the guarantee-maximizing $q(x)$ turns out to be increasing in x_i.

Substituting these conjectures into the SVO, we get

$$\lambda(\theta, x) = g(q(x_1 + x_2) + q'(x_1 + x_2)) - \frac{(h - g)q(x_1 + x_2)}{x_1 + x_2}.$$

We see that $\lambda(\theta, x)$ is actually independent of θ, so we set $L_\theta = L$. The complementary slackness conditions discussed in the previous paragraph then implies

$$\begin{aligned}
&g(q(x_1 + x_2) + q'(x_1 + x_2)) - \frac{(h - g)q(x_1 + x_2)}{x_1 + x_2} = L,\ x_1 + x_2 \leqslant \overline{y}, \\
&q(0) = 0, \quad q(x_1 + x_2) = 1,\ x_1 + x_2 \geqslant \overline{y}, \\
&g - \frac{(h - g)}{\overline{y}} \geqslant L.
\end{aligned} \tag{42}$$

The first two equations are self-explanatory. When $x_1 + x_2 > L$, $\lambda(\theta, x) = g - \frac{h-g}{x_1+x_2}$, so the third equation is necessary and sufficient for $\lambda(\theta, x) \geqslant L$ for all $x_1 + x_2 > L$. Since $q(y) \leqslant 1 = q(\overline{y})$, we must have $\lim_{y \nearrow \overline{y}} q'(y) \geqslant 0$, and hence the first and third equation imply

$$g - \frac{(h-g)}{\overline{y}} = L. \tag{43}$$

Then we can solve (42) as:

$$\overline{q}(y) = \begin{cases} \frac{L}{g} e^{-y} y^{(h-g)/g} \int_{z=0}^{y} e^z z^{-(h-g)/g} dz & y \leqslant \overline{y}, \\ 1 & y > \overline{y}, \end{cases} \tag{44}$$

where

$$L = \frac{g e^{\overline{y}} \overline{y}^{-(h-g)/g}}{\int_{z=0}^{y} e^z z^{-(h-g)/g} dz}. \tag{45}$$

Equations (45), (44) and (45) fully determine the mechanism $\overline{M} = (\mathbb{R}_+^N, \overline{q}, \overline{t})$, where $\overline{t}(x) = \overline{q}(x_1 + x_2) \left(g + (h-g) \frac{x_2}{x_1+x_2} \right)$. Then we have

$$G(\overline{M}) \geqslant L = \frac{g e^{\overline{y}} \overline{y}^{-(h-g)/g}}{\int_{z=0}^{y} e^z z^{-(h-g)/g} dz}.$$

Since L does not depend on the state θ, this lower bound on guarantee is independent of the prior μ. When both states are equiprobable, $\overline{M}$ forms a saddle point with $\overline{I}_\epsilon$ as $\epsilon \to 0$, since the lower bound on guarantee coincides with the upper bound on potential in (41).

3.7 The Literature

We conclude this section with a brief discussion of the literature. The Bayesian mechanism design problem and the revelation principle goes back to Myerson (1981) or earlier. Bergemann and Morris (2013, 2016) introduced the notion of Bayes correlated equilibrium of a game for characterizing possible equilibrium outcomes across all information structures. Bergemann et al. (2017) applied this methodology to solve for the revenue guarantee of the first-price auction.

Du (2018) pioneered the use of the robust-predictions methodology for mechanism design: in an auction setting with pure common values, Du (2018) constructs a sequence of mechanisms, indexed by the number of bidders, that extracts all of the surplus in the limit as the number of bidders goes to infinity. Importantly, Du (2018) introduced a version of the lower bounding program $\underline{G}(X(k^2))$ to prove the main result of that paper. Aside from the case of a single bidder, the mechanisms constructed by Du (2018) do not maximize the guarantee for a fixed number of bidders.

Subsequently, Bergemann, Brooks, and Morris (2016) constructed a saddle point for the special case of the common value auction with two bidders whose values are either 0 or 1, using versions of both bounding programs. This analysis was then generalized by Brooks and Du (2021b), who solved for revenue-guarantee-maximizing mechanisms in common value auctions, with an arbitrary number of bidders and value distribution. A special case of that solution was described above.

Brooks and Du (2024) presented versions of the bounding programs for general mechanism design problems, and proved that there is no duality gap for a class of auction problems. Brooks and Du (2025) introduced the particular notion of dual reduction described above, as a rigorous foundation for the bounding programs.[9]

Brooks, Du, and Feffer (2025) solve for revenue-guarantee-maximizing auctions when each agent's expected value is known, but the joint distribution of values and the information structure are unknown. Brooks, Du, and Zhang (2024b) construct binary action mechanisms for selling a large number of goods to a large number of agents. These mechanisms extract all of the surplus as the number of agents grows large. Brooks and Du (2023) constructed saddle points for a public expenditure problem, which for two agents can be reinterpreted as the bilateral trade problem described in the preceding section. Brooks, Du, and Haberman (2024a) generalize the theory of Brooks and Du (2024) to include restricted classes of information structures.

4 Other Approaches to Robustness in Mechanism Design

Economists and game theorists have long been concerned with issues of robustness in mechanism design. The standard approach has been critiqued in various ways. For recent comprehensive surveys of this literature, see Bergemann and Morris (2012) and Carroll (2019). We will focus our discussion on those threads in the literature that relate most closely to the informationally-robust approach that has been our focus.

4.1 Beliefs and higher-order beliefs

To frame the discussion, it is helpful to articulate a benchmark against which we may seek greater "robustness." The standard approach to Bayesian mechanism design is to take a single information structure as a complete and correct description of the environment.[10] By and large, it is also assumed that this information structure satisfies the common prior assumption. Moreover, for any given mechanism, if there are multiple equilibria, we suppose that the designer can coordinate the agents on the equilibrium that they most prefer. The known information structure, favorable equilibrium selection, as well with the assumption that the designer has the ability to implement a rich set of mechanisms, together imply the *revelation principle*: the designer can without loss restrict attention to mechanisms

[9]Myerson (1981) previously introduced a related but distinct notion of dual reduction for complete information games. See Brooks and Du (2025) for a detailed discussion.

[10]Even though we treat partial implementation as the benchmark, we do not mean to suggest that it preceded chronologically the other approaches to mechanism design that we refer to below.

in which each agent's action is a report of their private information, and in equilibrium, agents report truthfully (Myerson, 1981, 1986).

This approach to mechanism design has been criticized on various grounds. First and foremost, the standard model assumes a great deal of common knowledge among the agents and the designer, about both the information structure and what equilibrium is being played. With respect to the information structure, mechanism design is often conducted in highly stylized models, such as independent private values, which we have no reason to think are correct descriptions of information in the real world. The common knowledge assumption seems especially controversial with regard to agents' higher-order beliefs about one another's' information. A canonical reference for this critique is Wilson (1987), who argued that mechanisms should be "detail free."

One way to sidestep the problems arising from misspecification of the higher-order beliefs, either on the part of the designer or the agents, is to use mechanisms that have equilibria in dominant strategies (in the case where each agent knows their own preferences, i.e., private values) or in ex post equilibrium (in the case where there is interdependence in preferences). Dominant strategy and ex post implementation have been widely adopted in mechanism design as methods of achieving "robust" implementation and as a resolution of the Wilson critique. Indeed, recent work has sought to strengthen the implementation concept even further (Li, 2017).

A long line of research has explored foundations for dominant strategy and ex post implementation. Dasgupta, Hammond, and Maskin (1979) argued that dominant strategy implementation is equivalent to implementation regardless of agents' higher-order beliefs in private value environments. More recently, Bergemann and Morris (2005) studied foundations for ex post equilibrium. They distinguished between agents' possible "payoff" types, about which there is common knowledge, and "belief" types that are payoff irrelevant but parametrize subjective beliefs. They gave a set of sufficient conditions on preferences of the agents and the social choice correspondence (termed "separability") for which implementation for all belief types is equivalent to ex post implementation, where "ex post" is with respect to the realized payoff types. Bergemann and Morris (2005) also give examples where separability fails and a given social choice correspondence is implementable, but not with ex post incentive compatible mechanisms. Going further, Jehiel et al. (2006) and subsequent papers have argued that ex post implementation is a demanding concept, and that generically, the only social choice correspondences that can be implemented ex post are those that are constant.

The literature on ex post implementation relaxes common knowledge of higher-order beliefs but maintains common knowledge of the set of possible payoff types. In the case of Dasgupta, Hammond, and Maskin (1979), these payoff types are the possible ex post preferences of the agents (which they are assumed to know). In Bergemann and Morris (2005), the relationship between the payoff types and each agent's ex post preferences may be more involved. Moreover, the payoff types are presumed to capture everything about the agents' preferences that are payoff relevant to the designer.

In a somewhat different take, Chung and Ely (2007) studied a correlated private value auctions problem. Their exercise is the following: Fix a distribution over the agents' values for a good, and consider all of the (possibly non-common prior) information structures in which agents know their values. Suppose the seller evaluates a given mechanism by the worst-case across such information structures of the best equilibrium. The seller has the option of using a mechanism in which agents report their private values and for which truthtelling is a dominant strategy. For certain "regular" value distributions, Chung and Ely construct an information structure in which the seller can do no better than the best dominant strategy mechanism. In their terminology, this provides a "maxmin" foundation for dominant strategy mechanisms. Yamashita and Zhu (2018) and Chen and Li (2018) extended Chung and Ely's result to settings with interdependent values and other mechanism design problems. These papers also give examples where regularity fails, and the designer can do strictly better than with dominant strategy mechanisms. Moreover, the worst-case information structures always violate the common prior assumption.

Like Chung and Ely (2007), the informationally-robust approach surveyed in this article endows the designer with an expected utility preference. Also like Chung and Ely (2007), participation constraints play a central role in the informationally-robust theory. In contrast to both of these strands of the literature, the informationally-robust approach imposes the common prior and does *not* assume that the designer can select their preferred equilibrium. We will comment further on equilibrium selection below. But regarding the common prior assumption, whether this is a feature or a bug of the theory depends on one's view of the strength of forces in the world towards common knowledge, and whether the particular non-common prior beliefs that would be a worst-case outcome are themselves plausible.

Perhaps most importantly, relative to all of the aforementioned papers on ex post and dominant strategy implementation, the informationally-robust approach does not assume common knowledge of the component of the agents' private information that is payoff-relevant to the designer and the agents. In particular, we model the fundamental uncertainty as being about payoff-relevant *states*, as opposed to payoff *types*. The distinction is important: In the informationally robust approach, the designer is not presumed to place any restrictions on the agents' payoff-relevant information, and one possibility is that they know nothing at all. Hence, agents' information about what is payoff relevant cannot be disentangled from their information about others' information, and there is no set of payoff types with respect to which incentives could be provided ex post. Indeed, the only outcomes which could be implemented ex post are those that do not depend on private information at all. Thus, the designer is forced to rely on Bayesian mechanisms to achieve their objectives.

4.2 Equilibrium Selection and Strong Nash Implementation

A given mechanism and information structure can have many equilibria. The dominant paradigm in Bayesian mechanism design is to suppose that the equilibrium played will be the one that is most preferred by the designer. If truthful equilibria were not selected, then the revelation principle would not apply, and we would have to grapple with optimization over all indirect mechanisms and the whole set of Bayes Nash equilibria, a decidedly daunting task.

In contrast, the literature on mechanism design under complete information has predominantly concerned itself with achieving desirable outcomes in *all* equilibria (Maskin, 1999). This is so-called *strong* (or *full*) *implementation*, to be contrasted with *weak* (or *partial*) *implementation* when the designer picks the equilibrium. Strong implementation has also been studied in a Bayesian setting by Serrano and Vohra (2010). The mechanisms that achieve strong implementation often have features such as integer games (or modulo games when restricting to pure equilibria) for "killing off" undesirable equilibria. Relatedly, Abreu and Matsushima (1992b) study virtual implementation in mechanisms that have a unique strategy profile that survives iterated deletion of strictly dominated strategies. Their mechanisms divide the outcome into a series of small probability events, eliciting separate reports for each event in the sequence, and targeting a (small) punishment at the agent who is the "first" to disagree with the others' reports.

The informationally robust approach described in this article embraces elements of both traditions. In the potential, we consider the best equilibrium for the designer, but in the guarantee, we consider the worst equilibrium. Thus, when the min potential is equal to the max guarantee, we know that the designer achieves the same payoff regardless of whether they can select the equilibrium, and in particular, the mechanisms that maximize the expected lowest strategic virtual objective achieve the max guarantee in all equilibria. Moreover, this guarantee is achieved without resorting to integer games (or modulo games with restrictions on strategies) or with the targeted punishments and fine probabilistic structure of Abreu-Matsushima mechanisms.[11]

Is equilibrium multiplicity of significant practical concern, or is it more of a theoretical nuisance? Consider for example the second-price auction. In the independent private value model, each agent has a unique weakly undominated strategy to bid their value (Vickrey, 1961). But there are also "bidding ring" equilibria in which one agent makes a high bid, and the others essentially refuse to participate. By contrast, the payoff-equivalent equilibrium of the first-price auction is essentially unique (Lizzeri and Persico, 2000). Are the bidding ring equilibria plausible? Rothkopf, Teisberg, and Kahn (1990) have argued that they represent a realistic method of collusion in repeated auctions, wherein the agents take turns as the high bidder, and that second-price auctions are more vulnerable to collusion than first-price auctions as a result. At the same time, the designer could always "perturb" the mechanism by adding a noisy hidden reserve price, so that bidding one's value becomes strictly dominant. Moreover, this can be done with very arbitrarily small probability, and at negligible cost to the seller. Would such perturbations be effective in deterring collusive behavior? The answer has to depend on how large is the perturbation and how sensitive the bidders are to small changes in their payoffs.

While we are not aware of general results along these lines, it may be that for many mechanism design problems, the optimal payoff to the designer does not depend on what we assume about equilibrium selection, at least in the benchmark setting where the agents are sensitive to arbitrarily small changes in their payoffs. The reason would be because a

[11]The Abreu-Matsushima mechanisms have been criticized as artificial and implausible. See Glazer and Rosenthal (1992) for such a critique and Abreu and Matsushima (1992a) for a response. This view has recently been contested by Kapon, Del Carpio, and Chassang (2024), who view a variant of the Abreu-Matsushima targeted punishments as a practical tool for mechanism design in large populations.

mechanism that weakly implements the desired outcome can always be perturbed at low cost in a way that selects for the most preferred equilibrium. At the same time, there may be more than one way of selecting favorable equilibria. In a similar vein, there may be mechanisms that maximize the guarantee even though they do not maximize the expected lowest strategic virtual objective, e.g., by effectively asking the agents to report the information structure. Which class of solutions are preferred may come down to factors outside of the model, such as how sensitive we think real agents would be to the particular incentives provided by the mechanism. With regard to mechanisms that maximize the expected lowest strategic virtual objective, their performance is entirely driven by the local equilibrium constraints, that each agent prefers their equilibrium action to a marginal movement away from the secure action. This may be a more appealing method of eliminating undesirable equilibria than other devices that have been suggested in the literature.

4.3 Robustness to Fundamentals

In our definition of the guarantee in Section 2, we fixed the distribution over the payoff-relevant state θ, and took a worst-case over common-prior information structures with the given marginal. Thus, the "robustness" in the guarantee is with respect to just the agents' information and equilibrium, and not with respect to fundamentals of the economy. There is a large body of other work that emphasizes instead robustness with respect to the fundamentals themselves, and deemphasizes the role of information and equilibrium.

For example, Carroll (2017) studied a class of single agent screening problems where the payoff fundamental and outcome are multidimensional and there is additive separability in preferences across the dimensions. A leading example would be multi-product monopoly with additive values. A key finding is that there are correlation structures for which it is without loss to screen separately. Che and Zhong (2021) and Deb and Roesler (2023) further investigate the role of correlation in multiproduct monopoly and find cases where it is optimal to either sell separately or only offer a grand bundle consisting of all of the goods. And while certain moments of the value distribution are known, e.g., the marginal distributions, the correlation structure of not known. In multi-agent settings, He and Li (2022) analyzes guarantees of second-price auctions with (possibly random) reserve prices when the marginal distribution of each agent's private value is known but the correlation structure is uncertain. Che (2020) studies a similar environment and argues that the second-price auction with random reserve provides the highest guarantee from a class of "competitive" mechanisms. Zhang (2022) conducts a related exercise for bilateral trade and argues that double auctions with a random trade price maximize the guarantee among dominant strategy and ex post individually rational mechanisms.

The spirit of these exercises is somewhat different in that they are either consider a single agent with general mechanisms, or multiple agents with private values and dominant strategy mechanisms. In either case, higher-order beliefs are irrelevant to the analysis. While the informationally robust approach applies also to single-agent models, the results are most powerful in multi-agent Bayesian mechanism design, where higher-order beliefs play an essential role.

There is also a voluminous literature, predominantly in computer science, on algorithmic mechanism design that considers how the performance of a mechanism will vary across environments. The design criterion is the minimum ratio between a mechanism's performance and a given benchmark. As far as we are aware, the vast majority of work in this literature fixes a form for information (e.g., private values in auctions) and then considers robustness with respect to fundamentals. See Hartline (2012) for a discussion and Nisan et al. (2007) for a textbook treatment.

Of course, one could ask for robustness with respect to fundamentals and information. Brooks (2013) conducts such an analysis of private value auctions where only certain moments of the value distribution are known, and finds that a modified second-price auction maximizes a guarantee over all private value information structures satisfying a given moment condition. Within the informationally robust framework that has been the focus of this article, there is conceptual challenge associated with also entertaining uncertainty over the prior distribution of the payoff-relevant fundamental, which we denoted μ. Indeed, Brooks and Du (2021b, 2024) discuss how the performance of a guarantee maximizer at μ would vary as the prior changes to μ', and argues that there is a lower bound on the mechanism's guarantee that is linear in $\mu - \mu'$. Also, as discussed in Brooks and Du (2024, 2023), the results on the informationally robust model of bilateral trade that we reviewed in Section 3 remain true even if the designer only knows a lower bound on the gains from trade. Similarly, in Brooks and Du (2021a), we consider informationally robust optimal auction design when the designer only knows the expectation of each agent's value.

In our view, for mechanism design to be useful as a normative theory, the parameters of the model must be quantities that we could reasonably expect a user of the theory to be able to specify, either through empirical or introspective analysis. The worst-case criterion can be a useful modeling device for "solving out" those parameters that the designer is unable to quantify but on which the theory depends. For that reason, it may be advantageous to consider models in which the designer does not fully specify the prior distribution over the agents' preferences (parametrized by θ in the model of Section 2) but rather considers a set of priors that reflect the designer's knowledge of the environment.

5 A Case Against the Worst Case

The theory exposited in this article is one in which a designer plays their own devil's advocate: For each possible mechanism, the designer thoroughly explores its potential flaws and deficiencies. The hope is that such critical analysis will lead to a more robust solution to the problem of institutional design. After having considered and illustrated the potential benefits of this theory, it seems only fitting that we should similarly turn a critical eye on the theory itself, and assess the ways in which it may fail to achieve our stated goals. At the very least, we should identify those areas in which the theory requires further development.

5.1 The Worst-Case is not the Relevant Case

Perhaps the most obvious critique of this theory is evaluating a mechanism by its the worst-case outcome is too extreme, and may not be representative of the mechanism's performance in the most important environments. For example, suppose it were the case that the guarantee is maximized by a mechanism that achieves a payoff of 1 in all mechanisms and equilibria. There is another mechanism, however, that achieves a payoff of 2 in every information structure and equilibrium, except for a single information structure and equilibrium, in which it achieves a payoff of zero. Guarantee maximization would select the first mechanism. But the second mechanism might be preferred, if the particulars of the bad information structure/equilibrium made it a poor reflection of the environments in which we expect the mechanism to operate.

We should not think that a mechanism is "good" just because it maximizes the guarantee. Rather, we have to look at the particular mechanisms that arise from the theory, and assess whether the structures driving their performance would be effective with actual human beings as agents, and not just their theoretical representations. In addition, we should also look at those worst-case environments that the mechanisms are "guarding" against; if those environments seem plausible, then that is a further argument that the guarantee-maximizing mechanism is an efficacious design. Guarantee maximization in and of itself is a heuristic that we hope may guide us to novel and efficacious designs for mechanisms.

For example, in the case of proportional auctions, the mechanism distributes the allocation across agents in such a manner that both the transfer and the net sensitivity of payoffs to actions is independent of which agent is taking which action. This is a sensible structure if we are worried that agents might coordinate on who is taking which action in order to maximize their payoffs, to the detriment of the designer (as in bidding ring equilibria of the second-price auction) or if we are worried that the agents will load all of the allocation on the agent with the most valuable information, thus generating a strong winner's curse (as is possible in the first-price auction). Moreover, these mechanisms are rationalized by the potential-minimizing information structure in which the agents have independent signals, and the value is an increasing function of the average signal, which is not an especially exotic functional form.

But suppose it is the case that the potential-minimizing information structure is one that is implausible. One way to proceed would be to restrict the set of information structures to exclude the potential minimizer (or information structures that share whatever properties that make it implausible). For example, in common value auctions when the good has to be allocated, the potential-minimizing information structure has the feature that the join of the agents' information perfectly reveals the value. Perhaps we think it is implausible that by pooling their information, the agents would always know the value *exactly.* Brooks, Du, and Haberman (2024a) suggest that in this situation, we could put an "upper bound" on the agents' information. They develop a general methodology and use it to compute robust predictions for revenue in auctions and for the computation of guarantee-maximizing mechanisms.

Alternatively, it could be that in the worst-case, the agents have too *little* information. Indeed, for many problems problems, such as taxation or voting, participation constraints of the agents' seem less relevant, since the government has powerful punishments at its disposal. If we drop participation constraints from the informationally-robust design problem, then the solution is trivial: the worst case is that the agents have no information, and the designer can do no better implement the ex ante optimal outcome. But perhaps this worst-case is implausible: while we may be uncertain about the agents' information, we may think that there is a lower bound on what they know.

Even with participation constraints, we may wish to put lower bounds on agents' information. We began this article with a paean of praise for the independent private value model of Vickrey (1961) and Myerson (1981), where agents are supposed to know their values. And we suggested that independent private values can be thought of as a worst case of sorts (relative to correlated private value models). And yet, the theory of informationally robust predictions has thus far failed to provide a foundation for independent private values. One view is that private values is an technically convenient but unnatural assumption, and we should not be basing mechanism design on it. However, one could argue that agents know more about their own values than about others'. One way to capture this is with a lower bound on the agents' information, where each agent knows their own value and may know more.

Bergemann and Morris (2016) actually do incorporate such a lower bounds on the agents' information in their formulation of BCE: There is a fixed baseline information structure, and the agents are assumed to observe their baseline signals, plus some additional information. The theory of informationally robust mechanism design with lower bounds on information is still being developed. However, preliminary steps have been taken by Brooks and Du (2025), who generalize dual reductions to the case where there is a baseline information structure. In this case, it is without loss to consider mechanisms and information structures in which actions or signals are *sequences* of baseline signals, and the binding equilibrium constraints correspond to deviating to a sequence that is one entry longer (in the case of mechanisms and guarantees) or one entry shorter (in the case of information and potentials). It remains to be seen where this theory will lead, and whether it will lead to useful insights when there are lower bounds on agents' private information.

5.2 Participation Security is Too Strong

It is clear that participation constraints play a central role in the theory. But participation constraints are modeled in two different ways: As participation security for guarantees, and as interim individual rationality for potentials (i.e., non-negative interim utilities). Participation security implies interim IR in any information structure and equilibrium, but the converse is not true.

For the examples given above with zero duality gap, the difference in how we model participation constraints is immaterial: because the min potential is equal to the max guarantee, the designer could not achieve a higher guarantee even if they could pick their preferred equilibrium subject to interim IR. However, this is not always the case. The following example is from Brooks and Du (2023, 2024): A public good can be produced at a cost of one dollar. There are two agents who may fund it. The good generates the same value θ to both agents, which is either 0 or a number in $(1/2, 1)$, with likelihoods ϵ

and $1-\epsilon$ respectively. In this setting, if an agent takes a secure action, their contribution must be zero, so that if only one agent's action is not secure, they must pay the whole cost of the public good. Hence, any participation-secure mechanism has an equilibrium in which both agents take the secure action, and the good is not produced. However, for any information structure, there is an incentive compatible and individually rational direct mechanism where the good is produced if and only if the interim expectation of θ (conditional on both agents' signals) is at least 1/2, with each agent paying half the cost. When ϵ goes to zero, the surplus generated by this mechanism converges to the ex post efficient surplus of $2\theta - 1$, uniformly across information structures.[12]

Thus, participation security is in general much stronger than the requirement that for every information structure there exists an equilibrium with non-negative interim utility. Further advances in such problems may depend on finding tractable and conceptually appealing ways of relaxing participation security.

5.3 Randomization

Even if we have a saddle point in which the worst-case information and equilibrium are plausible, we still must ask, could the guarantee maximizing mechanism actually be implemented? Both proportional auctions and proportional-price trading mechanisms exhibit *interior* outcomes that respond smoothly to the agents' actions. This structure is not entirely surprising. Guarantee maximizers tend to equalize the SVO across action profiles, and doing so requires interior allocations (similarly to how maxmin mixed strategies in zero-sum games may equalize payoffs across opponent's actions).

When goods are indivisible, interior allocations correspond to randomization. One may be skeptical of a designer's ability to commit to such randomization, especially with carefully calibrated probabilities. However, if the goods being allocated are divisible, e.g., bushels of wheat or shares of common stock, then we can interpret the interior allocation as dividing the good across agents, which seems less challenging to implement in practice.

One might ask, in addition to randomization within the mechanism, could the designer improve the guarantee by randomizing over the mechanism itself?[13] This seems to have been precluded in the model of Section 2. However, it turns out such randomization over mechanisms is always *equivalent* to randomization within the mechanism. Indeed, suppose the designer implemented a lottery over mechanisms. The designer could equivalently put probability one on this mechanism's strategic normal form, in which each agent's action is a strategy that says, for each outcome of the randomization, which action do they play. A similar comment applies to information: A lottery over information structures is equivalent to a single information structure with different common knowledge components.

Nonetheless, when goods are indivisible, guarantee-maximizing mechanisms may involve randomization. If this randomization is seen as problematic, then one way to proceed is

[12] The worst-case information for this mechanism would be one for which the interim expected has two point support, and where the lower value approaches 1/2 from below (so that the probability that agents will not fund the good is maximized). The probability placed on interim values below 1/2 must go to zero as $\epsilon \to 0$.

[13] This would be analogous to how randomization over contracts can produce strictly higher guarantees than with deterministic contracts (Kambhampati, 2023).

solve a restricted problem where we maximize the guarantee over a class of deterministic mechanisms. This is the approach taken by Bergemann, Brooks, and Morris (2019), who optimized the guarantee for revenue single-item auctions across all "standard" auctions, a class that includes first-price, second-price, and all-pay auctions, as well as convex combinations of these rules. They find that the first-price auction has the highest guarantee of any such mechanism. Dovetailing with our earlier discussion of restricted sets of information structures, Bergemann, Brooks, and Morris (2019) also consider a restricted exercise where they consider pure common values and compute a restricted guarantee where the information must be an affiliated values environment and the agents play the symmetric monotone pure strategy equilibrium of Milgrom and Weber (1982). With this restriction, the guarantee of the first-price auction is unchanged, but now first-price and second-price auctions have the same guarantee. At the revenue minimizing BCE, all "upward" obedience constraints bind. An interesting direction for future research is to explore the relationship between randomization and other patterns of binding equilibrium constraints, which correspond to different notions of the strategic virtual objective.

5.4 Simplicity and Portability?

We motivated the informationally robust model with the twin desires for simplicity and portability. The former arose out of a belief that if a mechanism is too complicated, then equilibrium becomes less compelling, and we are not confident in our predictions agents will behave in the mechanism. We argued that the search for the worst-case information would lead us to an ordered structure on signals that reflects the efficient use of information to amplify the agents' objections. Such structure is also reflected in the guarantee maximizing mechanisms. Moreover, guarantee-maximizing mechanisms will always be as or more effective in achieving the designer's goals, regardless of the environment, as they are in the potential-minimizing information structures. In that sense, guarantee-maximizing mechanisms exhibit a strong form of portability across information structures and equilibria.

One can push back against these claims. The mere fact that the binding equilibrium constraints are linearly ordered is not by itself a satisfying definition of simplicity. For one thing, this is a kind of simplicity from the analyst's perspective, but what really matters is whether it is easy for the agents to converge to equilibrium. But even analytical tractability is not assured. Potential minimizers and guarantee maximizers are solutions to linear programming problems, but the solution may be still be hard to describe analytically. For example, in optimal auctions, potential-minimizing information structures always have independent signals, we do not know the analytical form of the optimal interim expected value in general. Thus, it may still be that the theory is tractable only for certain specifications of preferences and priors.

Regarding portability, we have already remarked that the guarantee is a rather extreme notion. Indeed, we should be asking for portability across the particular environments that are empirically relevant, and even then, portability should be traded off against optimality in environments that are regarded as more likely.

Perhaps a more serious issue is that by failing to articulate exactly when and why we think that equilibrium may become less convincing, we have undermined the claim that mechanisms with high guarantees are portable. For, without a better understanding of how agents reason through a mechanism in a complex environment, how can we even be confident in the equilibrium prediction, or even in the premise that that agents would be willing to participate?[14] One could argue that because guarantee maximizers only rely on local equilibrium constraints, the forces that drive the guarantee should be robust to agents who can only discover locally optimal actions. Still, it seems like we have avoided a central issue, which is exactly what kinds of complexity will make it will be for agents to learn how to play a mechanism and converge to equilibrium.

Indeed, the whole informationally-robust approach may be viewed as a workaround for not having been able to properly define or quantify simplicity or portability, so that it could be directly incorporated into the design problem. Echoing one of our earlier sentiments, the best hope for the informationally-robust approach is that in spite of not having precisely articulated what we are after, it may lead us there anyway, and we will know it when we see it. At the very least, we hope that it will generate novel ideas for how to structure incentives. But it may be that the ultimate progress of mechanism design depends on more directly addressing the issue of how agents reason through a mechanism and converge in their behavior, so that we may design institutions with greater regard for the realities and peculiarities of human strategic interaction.

References

Abreu, D. and H. Matsushima (1992a): "A response to Glazer and Rosenthal," *Econometrica: Journal of the Econometric Society*, 1439–1442.

——— (1992b): "Virtual implementation in iteratively undominated strategies: complete information," *Econometrica: Journal of the Econometric Society*, 993–1008.

Akerlof, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 488–500.

Bergemann, D., B. Brooks, and S. Morris (2016): "Informationally Robust Optimal Auction Design," Tech. rep., Princeton University and the University of Chicago and Yale University, working paper.

——— (2017): "First-Price Auctions with General Information Structures: Implications for Bidding and Revenue," *Econometrica*, 85, 107–143.

——— (2019): "Revenue guarantee equivalence," *The American Economic Review*, 109, 1911–29.

Bergemann, D. and S. Morris (2005): "Robust mechanism design," *Econometrica*, 1771–1813.

[14] Other work in mechanism design has attempted to address the cognitive limitations of agents, such as Li (2017) and Nagel and Saitto (2024) in the context of dominant strategy implementation in private value environments.

——— (2012): *Robust mechanism design: The role of private information and higher order beliefs*, vol. 2, World Scientific.

——— (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

Börgers, T. (2017): "(No) Foundations of Dominant-Strategy Mechanisms: A Comment on Chung and Ely (2007)," *Review of Economic Design*, 1–10.

Brooks, B. (2013): "Surveying and Selling: Belief and Surplus Extraction in Auctions," Working paper.

Brooks, B. and S. Du (2021a): "Maxmin Auction Design with Known Expected Values," Tech. rep., The University of Chicago and University of California-San Diego, working paper.

——— (2021b): "Optimal auction design with common values: An informationally robust approach," *Econometrica*, 89, 1313–1360.

——— (2023): "Robust Mechanisms for the Financing of Public Goods," Tech. rep., The University of Chicago and University of California-San Diego, working paper.

——— (2024): "On the structure of informationally robust optimal mechanisms," *Econometrica*, 92, 1391–1438.

——— (2025): "Dual Reductions and the First-Order Approach in Informationally-Robust Auction Design," Tech. rep., The University of Chicago and University of California-San Diego, working paper.

Brooks, B., S. Du, and J. Feffer (2025): "Compound Proportional Auctions," Tech. rep., The University of Chicago, The University of California-San Diego, and Stanford University, working paper.

Brooks, B., S. Du, and A. Haberman (2024a): "Robust Predictions with Bounded Information," Tech. rep., The University of Chicago and University of California-San Diego and Stanford University, working paper.

Brooks, B., S. Du, and L. Zhang (2024b): "An Informationally Robust Model of Perfect Competition," Tech. rep., The University of Chicago and University of California-San Diego, working paper.

Bulow, J. and P. Klemperer (1996): "Auctions Versus Negotiations," *The American Economic Review*, 180–194.

Carroll, G. (2017): "Robustness and separation in multidimensional screening," *Econometrica*, 85, 453–488.

——— (2019): "Robustness in mechanism design and contracting," *Annual Review of Economics*, 11, 139–166.

Che, E. (2020): "Distributionally Robust Optimal Auction Design under Mean Constraints," *arXiv:1911.07103*.

Che, Y.-K. and W. Zhong (2021): "Robustly-Optimal Mechanism for Selling Multiple Goods," in *Proceedings of the 22nd ACM Conference on Economics and Computation*, 314–315.

Chen, Y.-C. and J. Li (2018): "Revisiting the Foundations of Dominant-Strategy Mechanisms," *Journal of Economic Theory*, 178, 294–317.

Chung, K.-S. and J. C. Ely (2007): "Foundations of Dominant-Strategy Mechanisms," *The Review of Economic Studies*, 74, 447–476.

Crémer, J. and R. P. McLean (1985): "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53, 345–361.

——— (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 1247–1257.

Dasgupta, P., P. Hammond, and E. Maskin (1979): "The implementation of social choice rules: Some general results on incentive compatibility," *The Review of Economic Studies*, 46, 185–216.

Deb, R. and A.-K. Roesler (2023): "Multi-Dimensional Screening: Buyer-Optimal Learning and Informational Robustness," *arXiv preprint arXiv:2105.12304*.

Du, S. (2018): "Robust mechanisms under common valuation," *Econometrica*, 86, 1569–1588.

Ely, J. C. and M. Peski (2006): "Hierarchies of belief and interim rationalizability," *Theoretical Economics*, 1, 19–65.

Engelbrecht-Wiggans, R., P. R. Milgrom, and R. J. Weber (1983): "Competitive Bidding and Proprietary Information," *Journal of Mathematical Economics*, 11, 161–169.

Glazer, J. and R. W. Rosenthal (1992): "A note on Abreu-Matsushima mechanisms," *Econometrica: Journal of the Econometric Society*, 1435–1438.

Harsanyi, J. C. (1967): "Games with Incomplete Information played by "Bayesian" players, I–III Part I. The basic model," *Management science*, 14, 159–182.

Hartline, J. D. (2012): "Approximation in mechanism design," *American Economic Review*, 102, 330–336.

He, W. and J. Li (2022): "Correlation-robust auction design," *Journal of Economic Theory*, 200, 105403.

Jackson, M. O. (1992): "Implementation in undominated strategies: A look at bounded mechanisms," *The Review of Economic Studies*, 59, 757–775.

Jehiel, P., M. Meyer-ter Vehn, B. Moldovanu, and W. R. Zame (2006): "The limits of ex post implementation," *Econometrica*, 74, 585–610.

Kambhampati, A. (2023): "Randomization is optimal in the robust principal-agent problem," *Journal of Economic Theory*, 207, 105585.

——— (2025): "Proper Robustness and the Efficiency of Monopoly Screening," Tech. rep., Working Paper. 1, 2.

Kapon, S., L. Del Carpio, and S. Chassang (2024): "Using divide-and-conquer to improve tax collection," *The Quarterly Journal of Economics*, 139, 2475–2523.

Li, S. (2017): "Obviously strategy-proof mechanisms," *American Economic Review*, 107, 3257–3287.

Liu, Q. (2009): "On redundant types and Bayesian formulation of incomplete information," *Journal of Economic Theory*, 144, 2115–2145.

Lizzeri, A. and N. Persico (2000): "Uniqueness and existence of equilibrium in auctions with a reserve price," *Games and Economic Behavior*, 30, 83–114.

Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *The Review of Economic Studies*, 66, 23–38.

McAfee, R. P., J. McMillan, and P. J. Reny (1989): "Extracting the Surplus in the Common-Value Auction," *Econometrica*, 57, 1451–1459.

McAfee, R. P. and P. J. Reny (1992): "Correlated Information and Mechanism Design," *Econometrica*, 60, 395–421.

Mertens, J. F. and S. Zamir (1985): "Formulation of Bayesian analysis for games with incomplete information," *International journal of game theory*, 14, 1–29.

Milgrom, P. R. and R. J. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 1089–1122.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

——— (1986): "Multistage Games with Communication," *Econometrica*, 323–358.

——— (1997): "Dual reduction and elementary games," *Games and Economic Behavior*, 21, 183–202.

Nagel, L. and R. Saitto (2024): "As-if Dominant Strategy Mechanisms," in *Proceedings of the 25th ACM Conference on Economics and Computation*, 44–44.

Nisan, N., T. Roughgarden, É. Tardos, and V. V. Vazirani (2007): *Algorithmic Game Theory*, Cambridge University Press.

Pavan, A., I. Segal, and J. Toikka (2014): "Dynamic mechanism design: A myersonian approach," *Econometrica*, 82, 601–653.

Rothkopf, M. H., T. J. Teisberg, and E. P. Kahn (1990): "Why are Vickrey auctions rare?" *Journal of Political Economy*, 98, 94–109.

Serrano, R. and R. Vohra (2010): "Multiplicity of mixed equilibria in mechanisms: A unified approach to exact and approximate implementation," *Journal of Mathematical Economics*, 46, 775–785.

Vickrey, W. (1961): "Counterspeculation, Auctions, and Competitive Sealed Tenders," *The Journal of Finance*, 16, 8–37.

Wilson, R. (1977): "A Bidding Model of Perfect Competition," *The Review of Economic Studies*, 511–518.

——— (1987): *Game-theoretic analyses of trading processes*, Cambridge University Press.

Yamashita, T. and S. Zhu (2018): "On the foundations of ex post incentive compatible mechanisms," *American Economic Journal: Microeconomics*.

Yang, F. (2024): "Nested bundling," Tech. rep., Stanford University, working paper.

Zhang, W. (2022): "Random Double Auction: A Robust Bilateral Trading Mechanism," *arXiv:2105.05427*.