Simplicity and Portability in Mechanism Design: A Case for (and Against) the Worst Case*

Benjamin Brooks Songzi Du

October 30, 2025

Abstract

A central issue for mechanism design is how to identify theoretical environments that will lead to practically useful insights about optimal mechanisms. For a mechanism to be practically useful, it must be *simple* enough so that agents can optimize their behavior. The mechanism must also be *portable*, in that it also performs well in environments other than the one it was optimized for. We argue that by focusing on the informational environments that are the most challenging for designer, will necessarily identify optimal mechanisms that are both simple and portable. We survey a recent literature that operationalizes this idea, and we compare it to other approaches in robust mechanism design. Finally, we offer an (adversarial) critique of worst-case analysis in mechanism design.

Keywords: Mechanism design, information design, maxmin, informational 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, Joel Sobel, Bob Wilson, and Linchen Zhang. This research has been supported by NSF #2215259 and #2215475.

1 A Case for the Worst Case

1.1 The Problem with Mechanism Design

Roger Myerson has told the following story. It was the late 1970s, and Roger was working revenue maximization in auctions. He had used the revelation principle to reduce the problem to a linear program, with the distribution of values as a parameter. The question remained: What do optimal auctions look like, and how do they depend on the value distribution? For inspiration, Roger took a computational approach. At his request, Northwestern gave him a $100 budget to buy computer time on a communal mainframe that could run the simplex algorithm. He coded up (on punch cards) three minimalist examples of the linear program, with two bidders and two values, and ran them through the number cruncher at 25 cents a pop.[1] 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, and private information constrained how much revenue could be generated. Of course, in the first two programs, the values were correlated, and in the third they were independently distributed. As they say, the rest is history. Roger developed a general and complete analysis for independent values. This led to a number of insights that hugely influenced the subsequent development of mechanism design and auction theory, including the first-order approach in mechanism design, 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," and with the seller sometimes paying money to the buyers (Myerson, 1981). To this day, economic theorists are still trying to make sense of the correlated case.

[1]The rest of Roger's $100 budget remained unspent.

In this article, we adopt a normative perspective on mechanism design. Our purpose is to solve practical problems related to the allocation of resources under uncertainty. We view the theory as an engine for discovering designs for institutions that would be effective at shaping the behavior of real human beings. Two important desiderata, which we return to later in this introduction, are that: a mechanism has to be sufficiently *simple*, so that the agents can understand it and reliably make decisions that advance their own interest; and it has to be *portable*, meaning that it will work well in varied environments. As the preceding anecdote illustrates, there are parametric assumptions under which Bayesian optimal mechanisms are quite plausible candidates for meeting these criteria. But for other parameters, the optimal designs seem unintuitive or of doubtful practical usefulness. The anecdote also suggests that, somehow, whether an optimal mechanism is practically useful is related to how "challenging" is the environment that rationalizes the mechanism as optimal.

This discussion is, so far, incredibly vague. In the following pages, we review a literature that has articulated a precise sense in which mechanism design in the most challenging environments will necessarily lead us to mechanisms that are attractive solutions to practical mechanism design problems, in that they are necessarily both simple and portable.

1.2 The Worst-Case Approach

Let us expand on what we mean by the "environment." An essential ingredient in any theory of mechanism design is the description of the preferences and private information held by the agents who may interact with the mechanism. In the formal treatment below, as in the literature, these objects are modeled jointly as an *information structure*. The information structure describes everything that every agent knows about preferences and what others know. It implicitly pins down all the fine details of higher-order beliefs. The independent and correlated private value models in auction theory are two examples of information structures.

In Bayesian mechanism design, we analyze mechanisms that are optimal for a given information structure. As the example of optimal auctions shows, seemingly similar information structures can rationalize radically different mechanisms as being optimal, some of which seem quite plausible as candidates for practical implementation while others do not. To make progress in mechanism design, we have to somehow identify the information structures that lead us to practically useful designs, while also reconciling the theory with those cases where the implications of optimality are less compelling.

For the most part, our discipline follows an ad hoc approach to specifying the information structure, using a combination of intuition, received wisdom, and inspiration from the institutional context. The profession's track record of identifying useful information structures is rather mixed.[2] The theory that is the subject of this article takes an alternative and more systematic approach: We first fix certain payoff relevant aspects of the environment that we think the designer has a good handle on, such as the distribution of ex post preferences of the agents and the designer. Subject to those constraints on fundamentals, we then *derive* an information structure that is the most challenging, in that it induces the lowest possible payoff for the designer.

"Most challenging" can be interpreted in two complementary ways. It could mean the environment that is most challenging for a given mechanism, in the sense that it attains the mechanism's *guarantee*: its lowest performance across all information structures and equilibria. One way of identifying mechanisms that perform the best in the most challenging environments is to solve for mechanisms that maximize the guarantee. Alternatively, we could ask which environments are most challenging for *all* mechanisms. For a given information structure, its *potential* is the highest performance that can be achieved across all mechanisms and equilibria. Then, the uniformly most challenging information structures

[2]For example, a benchmark model of information with interdependent values is the mineral rights model of Wilson (1977) and the affiliated values model of Milgrom and Weber (1982). But with mineral rights or strict affiliation, "very strange auctions" are optimal and extract all of the surplus (McAfee, McMillan, and Reny, 1989; McAfee and Reny, 1992). For another example, in bilateral trade, the benchmark model is one of independent private values (Myerson and Satterthwaite, 1983). In the exceptional case where values are uniformly distributed, a double auction is optimal, but otherwise the theory does not yield sharp conclusions about the form of optimal mechanisms.

are those that minimize the potential. At a very high level, potential minimization is akin to focusing on information structures with independent types in the optimal auctions problem.[3]

These two programs, guarantee-maximization and potential-minimization, are dual to one another in the sense that they reverse the order of choosing the mechanism and the information structure. An important complication is that equilibrium selection also flips between the two programs (designer-best for potential and designer-worst for guarantee).

1.3 What We Know So Far

A key finding of the literature is that guarantee-maximizing mechanisms and potential-minimizing information structures have a very particular structure. In fact, there always exist potential-minimizing information structures that resemble the regular independent private value model from optimal auctions in fundamental ways. Each agent's private information can be ordered in a manner that we interpret to as a "degree of opposition" to the designer. The significance of this order is in how it relates to subset of equilibrium and participation constraints that pin down the min potential: the most opposed type has to be willing to participate, and every other type has to not want to mimic the type that is *slightly* more opposed. In reference to the similarity with independent private values, we refer to such information structures as *regular*.[4]

Analogously, there always exist guarantee-maximizing mechanisms with structure that mirrors that of the potential-minimizing information. Actions can always be ordered in a manner that we interpret as a "degree of participation," and the constraints that pin down the max guarantee are that: the least participatory action is always better than non-participation (a condition termed *participation security*); and whatever action an agent is

[3] As we review below, independence of types emerges naturally as a feature of potential-minimizing information structures in optimal auctions. Private values, on the other hand, do not naturally emerge from the baseline model. In Section 5, we will describe a richer model, where agents are *assumed* to know at least their own values. In this case, the independent private value model is indeed a potential minimizer, but the correlated private value model is not.

[4] In private-value auctions, the types are identified with private values values, and most opposed type corresponds to that with with the lowest possible value. But in general, the types in the potential minimizer are abstract, and the order has meaning only in terms of the pattern of relevant equilibrium constraints.

taking, they do not want to deviate to slightly higher participation. In a reuse of terminology, we also refer to such mechanisms as being *regular*. Thus, a fundamental result is that there always exist potential minimizers and guarantee maximizers that are regular.

Regularity concerns the structure of incentives, but it does not pin down the functional forms of potential minimizers and guarantee maximizers, which depend on the assumed distribution of ex post preferences. As we explain below, because of regularity, the dual programs of guarantee-maximization and potential-minimization reduce to linear programs. These programs have been solved in closed form for a number of design problems, including optimal auctions, public goods provision, and bilateral trade. The novel solutions include proportional auctions, compound proportional auctions, market order mechanisms, proportional cost-sharing mechanisms, and proportional-price trading mechanisms. The functional forms of guarantee maximizers are broadly consistent with the interpretation of the action as a degree of participation. For example, in a proportional auction, higher actions are associated with higher allocations and higher payments. And in proportional-price trading mechanisms, higher actions lead to higher probability of trade but on worse terms. In the potential minimizers, higher types are associated with higher values for the good in the case of optimal auctions, or greater value from trade in the case of bilateral trade, consistent with our interpretation of the types as a degree of opposition to the designer.

Finally, while it is not always the case, for many specifications—including all optimal auctions problems—there is no *duality gap*: min potential is exactly equal to max guarantee. Thus, potential-minimizing information structures rationalize guarantee-maximizing mechanisms as optimal. This is valuable for certifying that we have found solutions, and it also sheds light on which environments guarantee maximizers are implicitly guarding against.

1.4 Simplicity, Portability, and Practical Mechanism Design

Why is all of this relevant for "practical mechanism design"? This comes back to the two qualities that a mechanism must have for it to be practically useful: simplicity and portability. We will explain each of these in turn.

Starting with portability: For the most part, mechanisms should be viewed as long-lived institutions. After all, it may take considerable resources to establish infrastructure for implementing the mechanism, to persuade the agents to participate, and also for the agents to converge on optimal behavior. Indeed, the fact that it is costly to change the mechanism is the very source of the mechanism designer's commitment power. An effective design should therefore perform well even as the environment varies across space and across time. This is what we mean when we say that a practically useful mechanism has to be portable.

Returning to the motivating example of auctions, part of the appeal of the independent private value model as a normative theory is that the mechanisms that it rationalizes as optimal—first- and second-price auctions with reserve prices—also perform reasonably well in other environments, and even in the presence of correlation or interdependence in values.[5] By contrast, the bets in the "very strange auctions" that are optimal under correlated values depend on the precise probabilistic assessments held by the agents about one another's values. Under different beliefs, reporting values truthfully may no longer an equilibrium. What would the new equilibrium be, and how would the auction perform? We are not aware of any sharp results on this question, but it is reasonable to fear the worst: The optimal bets can involve large payments to or from the bidders, and under alternative beliefs, it could be that the bidders would either not want to participate or that they would participate but the auction would run a deficit.

[5] See, e.g., Milgrom and Weber (1982) and Bergemann, Brooks, and Morris (2017, 2019). The independent private model has also been praised as a positive theory of why first-price and English auctions are commonly used. As far as we know, the guarantee maximizers that have been identified in the literature have not been implemented in practice.

Now, to a game theorist, there is a seductive argument that portability can always be obtained for free, in the following manner. Suppose the designer is unsure whether the information structure is of type A or B, but an implicit assumption is that the true information structure is common knowledge among the agents. Why not build a bigger mechanism where the agents simply report whether the information structure is A or B and then also report their corresponding types? (And of course, to enforce truthful reporting, we can punish them if their statements are inconsistent.) With such a mechanism, it does not matter if the information structure is unknown to the designer; it will be discovered "online" through the agents' behavior. In fact, why not pursue this idea to its logical conclusion, and build a grand mechanism on a information structure that contains all of the myriad possibilities within it, such as the universal type space of Mertens and Zamir (1985)? Surely the resulting mechanism will be as portable as we could ever want!

In normative mechanism design, the outcomes, actions, and the rules of the mechanism have a very literal interpretation: they are features of the institution that the designer will ultimately construct. That is not to say that the ability to implement a mechanism can be taken for granted, and indeed, to put the mechanism into practice, one may have to overcome challenges related to computation, commitment power, and so on. But suppose it is possible to build such a grand mechanism as suggested in the previous paragraph, for an auction say. The bidder walks into the office and sits down at the desk. The designer explains: "Look, all you have to do is tell me what you believe about the values, what you believe about what others believe, what you believe others believe about what others believe, and so on. I have already carefully crafted allocations and payments that make truthful reporting the best thing for you. It's fine if you want to just name an information structure and your type, and I'll compute all those beliefs for you. And don't worry, I'm explaining this to the others, and I'm sure they will report truthfully as well." No doubt, the expression on the hapless bidder's face is somewhere between vacant confusion and twisted horror.

Drama aside, our point is the following. The information structure is an artifice that, together with Nash equilibrium, reflects forces towards common knowledge and best responses that we think genuinely exist in our world. As a benchmark, these assumptions discipline our reasoning about the performance of a mechanism under different scenarios. But, in contrast to the mechanism, an information structure and equilibrium are at best a loose metaphor for the way that real human beings incorporate private information into their behavior. We must be wary not to take the abstraction too seriously, lest it lead us to mechanisms that are so complex, in language and in intended mode of interaction, that real human beings would be unable to reason through the mechanism's rules and relate their possible choices to their private experience. If that were to happen, then our carefully crafted incentive structures would utterly fail in their intended purpose of shaping behavior.

Thus, for mechanisms to be useful, they have to be relatively *simple*, so that agents can understand the rules and discover how to advance their self interest.[6] Returning to optimal auctions, therein lies the dilemma with the correlated-value information structure: For the mechanisms that it rationalizes as optimal, there is no way to resolve the tension between simplicity and portability. If we model a rich set of possible beliefs, the mechanism will be too complex for agents to reliably play it in a manner that resembles equilibrium. But if our model of beliefs is too specialized, then we lose portability.

Returning now to guarantee maximization: Of course, we do not know how to quantify simplicity and portability in a satisfactory manner. If we did, we would just build those desiderata into the designer's objective function, or add simplicity and portability constraints on which mechanisms are feasible. Nonetheless, the literature has shown that guarantee-maximizing mechanisms are necessarily both simple and portable.

Simplicity is in the sense of regularity: Guarantee maximizing mechanisms are built like a ladder, where the rungs are the actions, the lowest rung secures participation, and incentives

[6] Could it be a good thing for the designer if the agents do not know how to optimize? If that were the case, it is hard to predict what they might do. But a reasonable concern is that it would undermine confidence in the mechanism, to the extent that agents might refuse to participate altogether.

are engineered so that in equilibrium the agents are pressured to move upwards one rung at a time. Endogenously, this pressure to move upwards is efficacious because the outcome induced by higher actions is more favorable to the designer. In fact, since the arguments underlying the guarantee only depend on local equilibrium constraints, the guarantee will hold as long as the agents are local optimizers.[7]

Why would this relatively simple incentive structure emerge endogenously from focusing on the most challenging environments? When there is no duality gap, guarantee maximizing mechanisms are guarding against the potential-minimizing information structures, which are also simple in the sense of regularity: types are ordered, participation binds for a lowest type, and the relevant incentive constraints point from each type to the next lower one.[8] But why should potential minimizers be regular?

If there were no participation constraints, then the worst-case information is straightforward: the potential is minimized when the agents have no information at all, and the designer can do no better than the ex ante optimal outcome. But with participation constraints, the ex ante optimum may be worse than not participating for some agent and in some state of the world. Informing the agents of when that happens enables them to object to the mechanism and therefore constrains the designer. But information does not only serve to inform the agents of when they would want to object; it also informs them of how they can manipulate the mechanism to their advantage, by behaving as if they had different information. This threat of manipulation can force the designer to distort the outcome towards what would appease the agent in those states of the world when they are most opposed. Regular information structures represent the most potent way to lever up objections through manipulation in order to constrain the designer, because any gains from tailoring

[7] This notion of simplicity concerns the structure of incentives, rather than the functional form of the mapping between actions and outcomes. In the examples that have been solved, the simplicity in the incentive structure seems to translate into simple functional forms for the guarantee-maximizing mechanisms, though this need not always be the case.

[8] As we explain below, there is also a possibility of a single type for each agent that has no binding truthtelling or participation constraints whatsoever. Such a type cannot arise in an environment with revenue maximization as the objective and transferable utility (Brooks and Du, 2025).

the mechanism's outcome to manage one type's capacity for objection or manipulation are offset by greater exposure to manipulation by another type. This is the threat that regular mechanisms are designed to guard against.

But what about portability? There are many ways in which one could evaluate how a mechanism's performance varies across information structures and equilibria, and we will discuss alternative viewpoints later in the article. But one natural measure of portability is the amount by which a mechanism might underperform, relative to what it achieves in the benchmark environment in which it is optimal. When max guarantee equals min potential, the guarantee maximizer is optimal at the potential minimizer and its optimal performance is the max guarantee. Hence, the guarantee maximizer can *never* underperform relative to this benchmark. It is in that sense that guarantee maximizers are the most portable mechanisms.

This concludes our introduction. We next describe the general theory and the main results outlined in the preceding paragraphs. We will also show how the theory can be operationalized with two examples: common value auctions and bilateral trade. After expositing the theory, we will compare it with other approaches to robust mechanism design. The last section of the article is a critique and a discussion of what we see as the most promising avenues for further development of the theory.

2 A Model of Mechanism Design

There is a finite number of agents, indexed by $i = 1, \dots, N$. There is also a mechanism designer who controls an *outcome* $\omega \in \Omega$. The agents and the designer have state-dependent expected utility preferences over outcomes, where the state is $\theta \in \Theta$, and the utility index is $u_i(\omega, \theta)$ for agent i and $w(\omega, \theta)$ for the designer.

For most of the formal results we describe in this survey, we will suppose that the states, outcomes, and other sets we introduce are finite, but for informal results and applications we will allow infinite sets. We return to this issue after presenting the rest of the model.

The designer builds a *mechanism* by which the outcome will be determined. The mechanism consists of, for each agent i, a set A_i of actions that can be taken in the mechanism, and a mapping m that associates to each profile of actions $a = (a_1, \ldots, a_N)$ and outcome ω a likelihood $m(\omega|a)$ that the given outcome 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. Note that actions are themselves payoff irrelevant, but the agents and the designer care about how actions influence the outcome. Let $\mathcal{M}$ be the set of mechanisms.

Given a mechanism M, an action a_i is *secure* if $\sum_\omega u_i(\omega, \theta) m(\omega|a_i, a_{-i}) \geqslant 0$ for every θ and a_{-i}. A mechanism is *participation secure* if each agent has a secure action. Let $\mathcal{M}^0$ be the set of participation-secure mechanisms, which we assume is non-empty.

The purpose of the mechanism is to incorporate the agents' private information about θ into the choice of outcome ω. 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. An information structure is formally equivalent to a common prior type space in the sense of Harsanyi (1967). We may refer to agent i with signal s_i as being a "type" of the agent (as we did in the introduction).

For most of the theory, we hold fixed the marginal distribution on Θ, denoted by $\mu \in \Delta(\Theta)$. Thus, the Knightian uncertainty of the designer is only about the agents' information and not about payoff-relevant fundamentals. Let $\mathcal{I}$ be the set of information structures for which the marginal of σ on θ is μ. In Section 4, we discuss the natural extension with Knightian uncertainty about μ as well.

Given a mechanism and an information structure, a *(behavioral) strategy* for agent i associates to each signal s_i and action a_i a likelihood $b_i(a_i|s_i)$. A strategy profile $b =$

$(b_1, \ldots, b_N)$ is identified with the mapping from signal profiles to lotteries over action profiles $b(a|s) \equiv \prod_i b_i(a_i|s_i)$.

Given a mechanism M, an information structure I, and a strategy profile b, the ex ante expected utility for agent i and the designer are respectively

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

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, which is non-empty as action and signal sets are finite. 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.

Given an information structure I, its *potential* $P(I)$ is the highest payoff the designer can achieve across all mechanisms and equilibria that satisfy interim participation:

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

We denote the infimum potential across all information structures by $P^* \equiv \inf_{I \in \mathcal{I}} P(I)$. An information structure that attains the infimum is a *potential minimizer*.

Given a mechanism M, its *guarantee* $G(M)$ is the lowest payoff for the designer across all information structures and equilibria:

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

We denote the highest possible guarantee over participation-secure mechanisms as $G^* = \sup_{M \in \mathcal{M}^0} G(M)$. A mechanism that attains the supremum is a *guarantee maximizer*.

We have modeled participation constraints in two different ways: In defining the potential, we have used a standard interim participation constraint, whereas in defining the guarantee we have imposed participation security on the mechanism. If M is participation secure, then for any I, any equilibrium of (M, I) must satisfy interim participation constraints. If not, then some type of some agent could profitably deviate to a secure action. Thus, agents are always willing to participate in a participation secure mechanism, regardless of the information structure.[9]

An elementary observation is that $G(M) \leqslant P(I)$ for any participation-secure mechanism M and information structure I. The reason is that (M, I) has an equilibrium b that satisfies interim participation constraints, and hence $G(M) \leqslant W(M, I, b) \leqslant P(I)$. It follows immediately that $G^* \leqslant P^*$. If $G^* < P^*$, there is a *duality gap*. The theory is especially powerful when there is no duality gap, so that max guarantee equals 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 is a *saddle point*: a pair (M, I) for which $G(M) = P(I)$. For such a pair, M is necessarily a guarantee maximizer (since no mechanism can have a guarantee higher than $P(I)$), and I is necessarily a potential-minimizing information structure (since no information structure can have a potential less than $G(M)$).[10]

---

[9] We discuss relaxing participation security in Section 5. We also show via an example that participation security of M can be more restrictive than the condition that $\mathcal{E}^0(M, I)$ is non-empty for every I.

[10] We cannot appeal to a minimax theorem for the existence of a saddle point, for at least three reasons: First, given a pair (M, I), we use the worst equilibrium for the designer when computing the guarantee and the best equilibrium when computing the potential. Thus, the game between maximizer and minimizer is not actually zero-sum. Second, even if we fix an equilibrium selection rule, we do not know of any linear or convex structure on equilibrium welfare that would allow us to apply the standard minimax theorems. Third, with the restriction to finite mechanisms and information structures, P^* and G^* may only be attained in the limit as the number of actions and signals goes to infinity. Nonetheless, in the next section we will describe sufficient conditions for $P^* = G^*$, in which case approximate saddle points always exist, meaning that for arbitrarily small ϵ, we can find a pair (M, I) with $P(I) \geqslant G(M) - \epsilon$.

The restriction to finite mechanisms is both conceptually beneficial and analytically convenient. On the conceptual side, if the mechanism is infinite, then there could be information structures and strategy profiles for which best responses do not exist. This is more than a technical concern; an infinite mechanism's guarantee might hold only vacuously, because no equilibria exist on any information structures. Indeed, we will confront this issue directly in the optimal auctions application that we describe below. Even if equilibria do exist for some information structures, the guarantee might be implausibly high because equilibria do not exist on information structures with low potential that would be especially challenging for finite mechanisms.[11] Our view is that if a mechanism's guarantee can not be approached with a sequence of finite mechanisms, then both the guarantee and the incentive structure underlying it are suspect. Note that for a finite mechanism M, in computing $G(M)$ it is without loss to take the infimum over finite information structures, due to the revelation principle, so that a finite mechanism's guarantee holds also for infinite information structures. Thus, the substantive assumption is that the mechanism is finite. On the analytical side, finiteness allows us to apply the powerful and elementary theory of finite dimensional linear programming for the main results.[12]

Even so, in order to obtain exact saddle points, it may be necessary to allow infinite mechanisms and infinite information structures. It is straightforward to generalize the definitions of mechanisms and information structures so that each A_i and S_i is a measurable set and strategies are probability transition kernels. The definitions of Bayes Nash equilibrium, guarantees, and potentials also generalize in the obvious way. This is the approach adopted in Brooks and Du (2021b), and which we informally describe in the examples below.

[11]This issue 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.

[12]We do not object to infinite information structures on conceptual grounds. Rather, the restriction to finite information structures is more for analytical convenience, and to parallel the finiteness of the mechanisms for which we compute guarantees.

3 Informationally Robust Optimal Mechanism Design

We now survey 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 then present two examples for which we explicitly construct saddle points: revenue guarantees in auctions and total welfare guarantees in bilateral trade.

3.1 The Lagrangian Formulation

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

$$\begin{aligned} L^p(S, \sigma, m, \alpha, \beta) \equiv & \sum_{s,\theta,\omega} w(\omega, \theta) m(\omega|s) \sigma(s, \theta) \\ & + \sum_{i,s_i'} \sum_{s,\theta,\omega} \alpha_i(s_i'|s_i) u_i(\omega, \theta)(m(\omega|s) - m(\omega|s_i', s_{-i}))\sigma(s, \theta) \\ & + \sum_{i} \sum_{s,\theta,\omega} \beta_i(s_i) u_i(\omega, \theta) m(\omega|s) \sigma(s, \theta). \end{aligned} \tag{1}$$

The first line in (1) contains the objective that the designer wants to maximize over direct revelation mechanisms $m : S \to \Delta(\Omega)$; the second line contains the incentive compatibility (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

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

where minimization over α and β determines the binding IC and IR constraints.

Analogously, 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 (Bergemann and Morris, 2016), which are direct recommendation information structures, in which each agent's signal is a "suggested" action from a metaphorical mediator, and it is an equilibrium for the agents to obey their recommendations. The corresponding Lagrangian is:

$$L^g(A, m, \sigma, \alpha) \equiv \sum_{a,\theta,\omega} w(\omega, \theta)m(\omega|a)\sigma(a, \theta)$$
$$- \sum_{i,a_i'} \sum_{a,\theta,\omega} \alpha_i(a_i'|a_i)u_i(\omega, \theta)(m(\omega|a)\sigma(a, \theta) - m(\omega|a_i', a_{-i}))\sigma(a, \theta).$$

The first line contains the designer's objective and the second line contains the obedience constraints (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

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

where maximization over α determines the binding obedience constraints.

In summary, the minimum potential is given by

$$\inf_{I \in \mathcal{I}} P(I) = \inf_{(S,\sigma) \in \mathcal{I}} \min_{\alpha \geqslant 0, \beta \geqslant 0} \max_{m: S \to \Delta(\Omega)} L^p(S, \sigma, m, \alpha, \beta), \tag{2}$$

and the maximum guarantee is given by

$$\sup_{M \in \mathcal{M}^0} G(M) = \sup_{(A,m) \in \mathcal{M}^0} \max_{\alpha \geqslant 0} \min_{\sigma \in \Delta(A \times \Theta): \mathrm{marg}_\Theta \sigma = \mu} L^g(A, m, \sigma, \alpha). \tag{3}$$

We note that the programs (2) and (3) are trilinear, since σ, m and α are multiplied together in L^p and L^g. Moreover, while the two Lagrangians contain similar terms for the designer's objective and the truthtelling/obedience constraints, the latter appear with opposite signs in the two Lagrangians.

3.2 Dual Reductions

An important insight into the structure of the solutions to (2) and (3) 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. This is proved using a "dual reduction" procedure, inspired by Myerson (1997), that operates jointly on the multipliers and the mechanism or information structure.

Let us illustrate how dual reduction works for the guarantee maximization problem. Fix a participation-secure mechanism $M = (A, m)$ and multipliers $\alpha_i(a_i'|a_i)$ for L^g, not necessarily optimal. 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 $a_i \in A_i$. Brooks and Du (2025) call this C a *deviation flow*.

Dual reduction produces a new (reduced) mechanism $\widehat{M}$ and new multipliers $\widehat{\alpha}$ for which

$$\min_{\widehat{\sigma} \in \Delta(X \times \Theta): \text{marg}_{\Theta} \sigma = \mu} L^g(X, \widehat{m}, \widehat{\sigma}, \widehat{\alpha}(C)) \geqslant \min_{\sigma \in \Delta(A \times \Theta): \text{marg}_{\Theta} \sigma = \mu} L^g(A, m, \sigma, \alpha). \quad (4)$$

Here is how the construction works. The actions in the reduced mechanism are non-negative integers $X_i = \{0, 1, 2, \ldots\}$, where each integer $x_i \geqslant 0$ is associated with a particular mixed strategy in the original mechanism M, which we denote by $b_i(\cdot|x_i)$. When agent i plays the action $x_i = 0$, the mixture puts probability one on an action $0 \in A_i$ that is secure in M, that is, $b_i(0|0) = 1$. The remaining mixtures are defined recursively according to

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

The reduced mechanism is $\widehat{M} = (X, \widehat{m})$, where $X = \prod_i X_i$, and for each $\omega \in \Omega$ and $x \in X$,

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

Thus, in the reduced mechanism, agents choose mixtures in M, actions are drawn independently, and then M is run. Finally, the new multipliers are

$$\widehat{\alpha}_i(C)(x'_i|x_i) = \begin{cases} C & \text{if } x'_i = x_i + 1; \\ 0 & \text{if } x'_i \neq x_i + 1, \end{cases} \tag{5}$$

Now, for any $\widehat{\sigma} \in \Delta(X \times \Theta)$, we can define $\sigma \in \Delta(A \times \Theta)$ such that $\sigma(a, \theta) = \sum_x \widehat{\sigma}(x, \theta) \prod_i b_i(a_i|x_i)$; by a straightforward calculation, $L^g(X, \widehat{m}, \widehat{\sigma}, \widehat{\alpha}(C)) = L^g(A, m, \sigma, \alpha)$, which proves (4).

In effect, the multipliers α represent 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 (by switching from x_i to $x_i + 1$), so that the implied lower bound on the guarantee can only go up.

While $\widehat{M}$ is not a finite mechanism, we can restrict actions to $X(k) \equiv \prod_i X_i(k)$, where $X_i(k) \equiv \{0, 1, \ldots, k\}$. Denoting this restriction as $\widehat{m}^k$ and using the same multiplier as in (5), Brooks and Du (2025) show that

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

We conclude that the max guarantee may be computed using multipliers of the form (5):

$$\sup_{M \in \mathcal{M}^0} G(M) = \sup_{k,C} \sup_{(X(k),m) \in \mathcal{M}^0} \min_{\sigma \in \Delta(X(k) \times \Theta) : \text{marg}_\Theta \sigma = \mu} L^g(X(k), m, \sigma, \widehat{\alpha}(C)). \tag{6}$$

What this result means is that in solving for guarantee maximizers, 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 a degree of participation, where the secure action 0 represents a minimal level of participation that is just better than non-participation. The multipliers $\widehat{\alpha}$ show that the constraints that matter are those associated with slightly increasing one's participation. Moreover, it is without loss to take the Lagrange multiplier on the binding local obedience constraints to be constant.[13]

A guarantee maximizer that solves the program on the right-hand side of (6) is *regular*, in the sense described in the introduction: Actions are ordered, the lowest actions are secure, and the guarantee is implied by local "upward" equilibrium constraints. Of course, the optimal value of (6) may only be attained in the limit as $k \to \infty$, in which case the optimal value can be approximated with finite mechanisms that are approximately regular, in the sense that minimum welfare subject to local upward constraints can be made arbitrarily close to the guarantee.

There is an analogous dual reduction for information structures that shows it is without loss to use a signal space $\overline{X}_i(k) = \{0, 1, \ldots, k\} \cup \{\infty\}$ for some k, and multipliers

$$\tilde{\alpha}_i(C)(x_i'|x_i) = \begin{cases} C & \text{if } x_i' = x_i - 1; \\ 0 & \text{if } x_i' \neq x_i - 1, \end{cases} \quad \tilde{\beta}_i(C)(x_i) = \begin{cases} C & \text{if } x_i = 0; \\ 0 & \text{if } x_i \neq 0. \end{cases} \tag{7}$$

[13]It is important to note that the constant local multiplier may not be the optimal dual solution to the linear program for the dual reduction. Nonetheless, this choice of obedience multipliers in (5) gives a lower bound on the guarantee that is still higher than that of the original mechanism.

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

$$\inf_{I \in \mathcal{I}} P(I) = \inf_{k,C} \inf_{(\overline{X}(k),\sigma) \in \mathcal{I}} \max_{m:\overline{X}(k) \to \Delta(\Omega)} L^p(\overline{X}(k), \sigma, m, \tilde{\alpha}(C), \tilde{\beta}(C)). \tag{8}$$

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 a likelihood of deviating, 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 objecting and leaving the mechanism entirely. 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 the signal in the original mechanism. Brooks and Du (2025) show that in the dual reduction, the deviations in which signal $x_i > 0$ mimics the signal $x_i - 1$ (or leaves if $x_i = 0$) replicate the stochastic deviation (α, β) in the original information structure. But fewer outcomes are feasible for the dual reduction, since the mechanism can only depend on the departure time, not the original signals. Moreover, the effect on the potential is negligible if we pool together all finite departure times above a sufficiently large threshold, so that the dual reduction can be approximated with finite information structures.[14]

Brooks and Du (2025) refer to an information structure for which this pattern of binding constraints is optimal as *regular*: the ordered signal is interpreted as a degree of opposition to the designer, and the binding constraint is that agents might want to slightly overstate their opposition, or to leave the mechanism if they have the strongest opposition. This is the

[14] For an extreme example of dual reduction of information, consider the correlated private value model in which full surplus extraction is possible. In this case, no incentive constraints bind, and every type has a binding interim participation constraint. Thus, in the associated Markov chain, all agents leave immediately, so that in the dual reduction, the agents have no information. Surely, optimal revenue under no information is weakly less than the full surplus, and strictly so if the values are not common, since it will be infeasible to allocate the good efficiently if there is no information.

result to which we referred in the introduction, when we asserted that potential minimizers would necessarily be simple in the sense of regularity.[15]

3.3 The Bounding Programs

Thus, it is without loss to restrict attention to the simple pattern of equilibrium constraints according to (5) and (7), where all of the non-zero multipliers have the same value. As we noted above, the multiplier C on local obedience/truthtelling constraints can always be made larger. In the context of the dual reduction of a mechanism, this can also be interpreted as making the actions more similar to each other, since the rate of switching actions in original mechanism is inversely proportional to C.

For that reason, 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 labeled as $X_i(k^2) = \{0, 1/k, 2/k, \ldots, k\}$. As $k \to \infty$, $X_i(k^2)$ "fills in" all of $\mathbb{R}_+$, the non-negative real numbers.[16] Since the space between consecutive signal/action in $X_i(k^2)$ is $1/k$, we consider $C = k$ in (5) and (7), 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 limit mechanism or information structure.

Substituting (5) and (7) with $C = k$ into L^p and L^g yields:

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

[15]As with mechanisms, the regularity of potential minimizers may only be approximate, if the solution to the right-hand side of (8) is only attained in the limit as k goes to infinity. See Brooks and Du (2025) for a more detailed discussion.

[16]For simplicity of exposition, we have omitted the special signal realization for which all constraints are slack. As shown by Brooks and Du (2021b, 2024), dual reduction will never assign positive probability to the infinite signal in optimal auctions, although it can arise for certain formulations of the public good and bilateral trade problems described in Brooks and Du (2023).

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

where for a function $f : X(k^2) \to \mathbb{R}$, we define the discrete derivatives

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

Thus, the difference in directions in $\widehat{\alpha}$ and $\tilde{\alpha}$ aligns the signs on incentive terms 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 converge to differentiable functions of x.[17]

Since they are obtained by restricting the problems (2) and (3), $L^p(X(k^2),\sigma,m,k)$ and $L^g(X(k^2),\sigma,m,k)$ yield upper and lower bounds on the potential and guarantee, respectively:

$$\overline{P}(X(k^2),\sigma) \equiv \max_{m:X(k^2)\to\Delta(\Omega)} L^p(X(k^2),\sigma,m,k) \geqslant P(X(k^2),\sigma)$$
$$\underline{G}(X(k^2),m) \equiv \min_{\sigma\in\Delta(X(k^2)\times\Theta):\mathrm{marg}_\Theta\,\sigma=\mu} L^g(X(k^2),\sigma,m,k) \leqslant G(X(k^2),m),$$

Optimizing over σ and m, respectively, yields the following bounding programs:

$$\overline{P}(X(k^2)) = \min_{\sigma\in\Delta(X(k^2)\times\Theta):\mathrm{marg}_\Theta\,\sigma=\mu} \overline{P}(X(k^2),\sigma), \tag{9}$$
$$\underline{G}(X(k^2)) = \max_{m:X(k^2)\to\Delta(\Omega)} \underline{G}(X(k^2),m), \text{st:} \sum_\omega u_i(\omega,\theta)m(\omega|0,a_{-i}) \geqslant 0\ \forall i,\theta,a_{-i}, \tag{10}$$

[17] This statement is made precise in Proposition 6 of Brooks and Du (2024).

with

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

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 (9) 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 (10) 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, which 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.

Returning to the general case, we can rewrite the objective in (10) as

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

where

$$\lambda(x, \theta) \equiv \sum_{\omega} w(\omega, \theta) m(\omega|x) + \sum_i \sum_{\omega} u_i(\omega, \theta) \nabla_i^+ m(\omega|x). \tag{11}$$

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

$$\underline{G}(X(k^2),m) = \min_{\sigma\in\Delta(X(k)\times\Theta):\mathrm{marg}_\Theta \sigma=\mu} L^g(X(k^2),\sigma,m,k) = \sum_\theta \mu(\theta)\min_x \lambda(x,\theta). \tag{12}$$

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 increase participation.

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

$$\gamma(x,\omega) \equiv \sum_\theta w(\omega,\theta)\sigma(x,\theta) - \sum_i\sum_\theta u_i(\omega,\theta)\tilde{\nabla}_i^+\sigma(x,\theta), \tag{13}$$

and for a function $f : X(k^2) \to \mathbb{R}$,

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

We refer to $\gamma(x,\omega)$ as the *informational virtual objective* (IVO). Thus,

$$\overline{P}(X(k^2),\sigma) = \max_{m:X(k)\to\Delta(\Omega)} L^p(X(k^2),\sigma,m,k) = \sum_x \max_\omega \gamma(x,\omega). \tag{14}$$

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 opposition.

The SVO and the IVO play central roles in the engineering of saddle points, as we will see in the applications. 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 mechanisms and information structures. But as we transition to examples, it will be more convenient to work with the continuum analogues 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, 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 characterize solutions.

All of this sounds quite loose. 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 (11))

$$\lambda(x, \theta) \equiv \sum_{\omega} w(\omega, \theta) m(\omega|x) + \sum_{i} \sum_{\omega} u_i(\omega, \theta) \nabla_i m(\omega|x).$$

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 and participation for the lowest type are necessary conditions for equilibrium, weak duality implies that

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

where (cf. equation (13))

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

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

We can describe this approach to solving for maximum guarantees problem as "guess and verify," which can succeed only if a saddle point exists and there is no duality gap. This is to be contrasted with the more systematic method using 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.

Nonetheless, for problems in auctions, public goods, and bilateral trade, the guess and verify approach has yielded results. The key step is to generate informed guesses for the saddle point (M, I), which may then be verified via weak duality. One way to do so is to utilize the bounding programs (9) and (10), which are finite dimensional linear programs (cf. equations (12) and (14)).[18] As in Roger Myerson's anecdote from the beginning of this article, numerical solutions of (9) and (10) can be used to generate and test conjectures and to infer functional forms in the continuum limit.

[18]To convert (10) into a linear program, one can introduce auxiliary variables $\lambda(\theta) \in \mathbb{R}$ for each θ, and add linear constraints of the form $\lambda(\theta, a) \leqslant \lambda(\theta)$ for all a, θ. Maximizing $\sum_\theta \lambda(\theta)$ over $\lambda(\cdot)$ and $m(\cdot|\cdot)$ that are feasible for (10) and these additional constraints is a finite linear program whose value is the same as (10) and whose optimal m is a solution to (10). A similar approach can be taken to convert (9) into a linear program.

Another way to guess saddle points is using linear programming duality. As argued in Brooks and Du (2024), the bounding programs (9) and (10) are an "approximate" dual pair, except that they differ in the direction of equilibrium constraints and the form of the participation constraint. 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*[19]:

$$m(\omega|x) > 0 \Longrightarrow \omega \in \operatorname{argmax}_{\omega'} \gamma(x,\omega'), \tag{15}$$

$$\sigma(x,\theta) > 0 \Longrightarrow x \in \operatorname{argmin}_{x'} \lambda(x',\theta). \tag{16}$$

There is no formal result in this literature that establishes complementary slackness as either a necessary or sufficient condition for a saddle point. However, complementary slackness does hold for the saddle points that have been constructed, and it has proven valuable as a heuristic for engineering saddle points, as we now illustrate.

3.5 Example: A Common Value Auction

We now describe the common value auction problem of Brooks and Du (2021b). A single unit of a good is for sale, for which the common value is $\theta \in \Theta \subset \mathbb{R}$. The possible outcomes are $\Omega = \{0,1,\ldots,N\} \times \{-t_{max}, t_{max}\}^N$, and for $\omega = (\iota, \tau_1, \ldots, \tau_N)$ and θ, the utilities are

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

[19] 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 (9) and (10)). Then complementary slackness in the sense of (15) and (16) is equivalent to (σ,m) being a Nash equilibrium.

Thus, if $\iota \in \{1, 2, \ldots, N\}$, the good is sold to agent ι, and otherwise the good is not sold, and τ_i is agent i's transfer to the designer. The designer only cares about revenue.

While τ_i can only take on two values, randomizing over these values induces an expected payment that we interpret as a continuous transfer. Indeed, given a lottery over ω, let q_i be the probability that i is allocated the good and t_i is agent i's expected payment. Then agent i's expected utility is $\theta q_i - t_i$, and the designer's expected revenue is $\sum_i t_i$. Hence, we may model the outcome in reduced form as $q \in \Delta(\{0, 1, \ldots, N\})$ and $t \in [-t_{max}, t_{max}]^N$. In the following analysis, summation over i means for $i \in \{1, 2, \ldots, N\}$, e.g., $q_0 = 1 - \sum_i q_i$. We will eventually consider t_{max} to be relatively large, so that the constraint on the size of the transfer will not bind.[20]

3.5.1 Potential-minimizing information structure

To compute the min potential, we solve for the information structure of the form $(\mathbb{R}_+^N, \sigma)$ that minimizes the expected highest IVO. Plugging in functional forms, we have

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

This expression is linear in t_i, so the optimum is 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). Thus, the constraint on the size of the transfers would not bind only if the coefficient of t_i is exactly zero, that is,

$$\sum_\theta (\sigma(x, \theta) + \nabla_i \sigma(x, \theta)) = 0 \tag{17}$$

for all i and $x \in \mathbb{R}_+^N$. We will analyze solutions that satisfy (17), and then subsequently verify via weak duality that they minimize the potential when t_{max} is sufficiently large.

[20] Brooks and Du (2021b) formally consider the case where the transfer is unbounded. This is not subsumed within our model with finitely many outcomes. Brooks and Du (2024) derive analogues of the bounding programs when transfers are free variables.

Let $\rho(x) \equiv \sum_\theta \sigma(x,\theta)$ denote the marginal on signals, and rewrite (17) as $\rho(x) + \nabla_i \rho(x) = 0$. The unique solution to this differential equation is $\rho(x) = \rho(0, x_{-i})e^{-x_i}$, and by iterating across agents, we conclude $\rho(x) = \overline{\rho}(x) \equiv e^{-\Sigma x}$, where $\Sigma x \equiv \sum_i x_i$. In other words, the signals must be iid standard exponential.

Next, let $v(x)$ be the expected value conditional on the signal profile x, i.e.,

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

Then, taking $\rho = \overline{\rho}$, transfers drop out of the IVO, and the remaining terms are

$$\gamma(x,(q,t)) = -\sum_i \sum_\theta \nabla_i \sigma(x,\theta)\theta q_i = -\sum_i \nabla_i(\overline{\rho}(x)v(x))q_i = \overline{\rho}(x)\sum_i (v(x) - \nabla_i v(x))q_i,$$

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, $\gamma(x,(q,t))$ is maximized by concentrating q on agents with the maximal virtual value $v(x) - \nabla_i v(x)$, if the max is non-negative, and otherwise the seller keeps the good. This suggests that to minimize $\overline{P}(\mathbb{R}^N_+, \sigma)$, virtual values should be equalized across i. That is the case if the interim value function only depends on Σx (so that $\nabla_i v(x)$ is independent of i). Brooks and Du (2021b) provide a general solution of this form.

For the rest of the example, we focus on the special case where $\Theta = \{0,1\}$, in which case $v(x)$ can also be interpreted as the interim probability that the value is equal to 1. Brooks and Du (2021b) show that the solution has the form $v(x) = \min\{e^{\Sigma x - \overline{y}}, 1\}$, for a constant $\overline{y} > 0$ that is chosen so that the ex ante probability of the low value matches the prior. When $\Sigma x \leqslant \overline{y}$, the IVO is exactly zero, so that the seller is also indifferent to keeping the good.

As a sum of iid exponentials, the random variable $y = \Sigma x$ follows an Erlang distribution, whose density and cumulative distribution are respectively $g_N(y)$ and $G_N(y)$. Thus, in order for for the law of iterated expectations to hold, it must be that the expectation of $v(x)$ is equal to $\mu(1)$, that is, $\overline{y}$ solves $\int_{y=0}^{\infty} \exp(y - \overline{y}) g_N(y) dy = \mu(1)$. Since the virtual value is 0 below $\overline{y}$ and 1 above $\overline{y}$, the resulting expected highest IVO is simply $1 - G_N(\overline{y})$.

3.5.2 Guarantee-maximizing mechanism

To compute the max guarantee, we solve for a mechanism of the form $(\mathbb{R}_+^N, q, t)$ that maximizes the expected lowest strategic virtual objective, where $q : \mathbb{R}_+^N \to \Delta(\{0, 1, \ldots, N\})$ and $t : \mathbb{R}_+^N \to [-t_{max}, t_{max}]^N$ are the expected allocation and transfer rules. The SVO is:

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

We can guess the correct solution using complementary slackness (16): If a pair (x, θ) is in the support of the potential minimizer, then x minimizes the SVO at state θ. Given our candidate potential minimizer, we then conjecture that every x minimizes the SVO when $\theta = 1$, and any x with $\Sigma x \leqslant \overline{y}$ minimizes the SVO when $\theta = 0$.

Pursuing this conjecture, let suppose there are constants L_0 and L_1 such that $L_1 \equiv \lambda(1, x)$ for all x, $L_0 \equiv \lambda(0, x)$ when $\Sigma x \leqslant \overline{y}$, and $L(0, x) \geqslant L_0$ when $\Sigma x > \overline{y}$. This helps us pin down the functional form of the allocation, since the divergence of q must satisfy

$$\nabla \cdot q(x) \equiv \sum_i \nabla_i q_i(x) = \lambda(1, x) - \lambda(0, x) \begin{cases} = L_1 - L_0 & \text{if } \Sigma x \leqslant \overline{y}; \\ \leqslant L_1 - L_0 & \text{if } \Sigma x > \overline{y}. \end{cases}$$

We can also use complementary slackness in the other direction, condition (15): When $\Sigma x > \overline{y}$, keeping the good does not maximize the IVO of the potential minimizer, and so $\sum_i q_i(x) = 1$. A solution to this PDE is

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

Thus, if $\Sigma x \leqslant \overline{y}$, agent i's allocation is $x_i/\overline{y}$. We can therefore interpret the action x_i as (up to a linear rescaling of units) a demand for a quantity of the good. If the demands exceed the supply, then the good is rationed proportionally. The resulting divergence is

$$\nabla \cdot \overline{q}(x) = \nabla \cdot \overline{q}(\Sigma x) = \begin{cases} N/\overline{y} = L_1 - L_0 & \text{if } \Sigma x \leqslant \overline{y}; \\ (N-1)/\Sigma x & \text{if } \Sigma x > \overline{y}. \end{cases} \tag{18}$$

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 unique solution.

Given the proportional form for the allocation, it is natural to conjecture that the transfer function should be proportional as well: for some total transfer function $T : \mathbb{R}_+ \to \mathbb{R}$,

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

Brooks and Du (2021b) call a mechanism $\overline{M} = (\mathbb{R}^N_+, \overline{q}, \overline{t})$ of this form a *proportional auction*.

Using the proportional form, we can simplify the SVO, and use the condition that $\lambda(1, x) = L_1$ to pin down the functional form for T:

$$\sum_i (\overline{t}_i(x) - \nabla_i \overline{t}_i(x)) = T(\Sigma x) - \frac{N-1}{\Sigma x} T(\Sigma x) - T'(\Sigma x) = L_1 - \nabla \cdot \overline{q}(\Sigma x).$$

The boundary condition $T(0) = 0$ is given by participation security. The solution to this differential equation is

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

where $g_N(y) = y^{N-1} e^{-y}/(N-1)!$, the density of the Erlang distribution, has now reappeared.

The resulting expected lowest SVO is

$$\sum_{\theta} \mu(\theta) \min_{x} \lambda(\theta, x) = \mu(0)L_0 + \mu(1)L_1 = L_1 - \mu_0 \frac{N}{\overline{y}}.$$

To maximize the lower bound, L_1 should be as large as possible. But if $L_1 > \int_{y=0}^{\infty} \nabla \cdot \overline{q}(y) g_N(y) dy$, then because $\lim_{y\to\infty} g_N(y) = 0$, we would have $\lim_{y\to\infty} T(y) = -\infty$. This would preclude the existence of any equilibrium, since the agents would race to bid an ever larger x_i, to extract money from the seller. Thus, the optimal value is to set L_1 exactly equal to $\int_{y=0}^{\infty} \nabla \cdot \overline{q}(y) g_N(y) dy$, in which case the aggregate transfer T will be bounded. Moreover, as long as $t_{max} \geqslant \sup_y T(y)$, the mechanism we have constructed is feasible, and the constraint on the size of the transfers does not bind.

With these optimal parameters, and using the functional form of g_N and the definition of $\overline{y}$, one can compute the expected lowest SVO:[21]

$$\int_{y=0}^{\infty} \nabla \cdot \overline{q}(y) g_N(y) dy - \mu(0)\frac{N}{\overline{y}} = G_N(\overline{y})\frac{N}{\overline{y}} + 1 - G_{N-1}(\overline{y}) - \mu(0)\frac{N}{\overline{y}} = 1 - G_N(\overline{y}) \qquad (19)$$

Thus, the expected lowest SVO of this proportional auction is *exactly* the expected highest IVO of the information structure we constructed previously. This proves that $P^* = G^* = 1 - G_N(\overline{y})$ and that $(\overline{M}, \overline{I})$ is a saddle point.

This is a good moment to remind the reader that while we used the complementary slackness heuristic to guess the functional form of the saddle point, the ultimate proof that this pair is a saddle point does not rely on complementary slackness; rather, it follows from the fact that the expected lowest SVO of $\overline{M}$ is equal to the expected highest IVO of $\overline{I}$.

Brooks and Du (2021b) show that, remarkably, the game $(\overline{M}, \overline{I})$ has an equilibrium in which each agent's action is equal to their signal. In other words, $\overline{M}$ is a revenue-maximizing direct revelation mechanism on $\overline{I}$, and $\overline{I}$ is a revenue-minimizing BCE (i.e., a

[21] Specifically, the integral evaluates to $\frac{N}{\overline{y}} \int_{y=0}^{\overline{y}} g_N(y)dy + \int_{y=\overline{y}}^{\infty} g_{N-1}(y)dy = \frac{N}{\overline{y}} G_{N+1}(\overline{y}) + 1 - G_N(\overline{y})$. In addition, it is straightforward to show that $\mu(1) = \int_{y=0}^{\infty} \min\{e^{y-\overline{y}}, 1\} g_N(y)dy = 1 - G_{N+1}(\overline{y})$.

direct recommendation information structure) on $\overline{M}$. Brooks and Du (2021b) refer to this phenomenon as a *double revelation principle*. The same property holds for the saddle point in the bilateral trade example we discuss next. Why this property should arise, and whether it is in fact a "principle" that holds more broadly, remains a mystery.

Brooks and Du (2021b) also show the striking fact that as $N \to \infty$, the optimal guarantee in (19) 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 strengthens an earlier result of Du (2018). Thus, the optimal guarantee converges to full surplus as the market gets large, regardless of how information and equilibrium changes as we add more agents to the market. Such an asymptotic full surplus extraction is not attained by a standard auction like the first price auction, as shown by Engelbrecht-Wiggans, Milgrom, and Weber (1983) and 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 a seller with a unit of a good; agent 2 is a potential buyer. The seller's value is equally likely to be 0 and h, and the buyer's value is $g > 0$ more than the seller's. Thus, the state θ is the ordered pair of seller and buyer values, and is equally likely to be $(0, g)$ and $(h, h + g)$.

The set of outcomes is $\Omega = \{0, 1\} \times \{-t_{max}, t_{max}\}$, with a typical element (ι, τ). The first coordinate indicates whether or not trade takes place, and the second coordinate is the net transfer from the buyer to the seller. As in the common value auction, we look at the case where t_{max} is large. The seller maximizes net surplus from the sale $\tau - \theta_1\iota$, the buyer maximizes net surplus from purchase $-\tau + \theta_2\iota$, and the designer maximizes expected gains from trade $\iota(\theta_2 - \theta_1) = \iota g$. Once again, denoting the probability of trade by q and the expected transfer by t, we can rewrite the preferences of seller, buyer, and designer in reduced form as $t - \theta_1 q$, $-t + \theta_2 q$, and $q(\theta_2 - \theta_1) = qg$ respectively.

We assume $h > g$ 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 private information.

3.6.1 Potential-minimizing information structure

Once again, we compute the min potential by solving for the information structure $(\mathbb{R}_+^N, \sigma)$ that minimizes the expected highest IVO, which in this case is

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

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 $x_1 + x_2$, which we now denote as $\rho(x_1 + x_2)$.

Now let $\nu(x) = \sigma(x,(h,h+g))/\rho(x_1+x_2)$ be the interim probability of the high state conditional on the signal. Substituting this back into the IVO and supposing that ρ and ν are differentiable, we have

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

Numerical simulations suggest that the optimal interim value function is of the form $\overline{\nu}(x) = x_2/(x_1+x_2)$. That is, conditional on the signal sum $x_1 + x_2$, a higher signal x_1 for the seller is a bad news about the value, while a higher signal x_2 for the buyer is a good news. Either way, a higher x_i is associated with agent i being more willing to trade at given terms. With this functional form, the IVO is once again a function of just $x_1 + x_2$:

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

The numerical simulations also suggest that, as in the common value auction, to minimize $\max_{q,t} \gamma(x, (q, t))$, we should pick ρ so that the resulting IVO is zero:

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

The solution is of the form $\rho(y) = Ae^y y^{-h/g}$ for some constant $A > 0$. However, unlike the information structure in the common value auction, we have $\lim_{y \to \infty} \rho(y) \neq 0$, so we must truncate this density at a finite upper bound. Moreover, if $h/g \geqslant 2$, $\overline{\rho}$ is not integrable around zero. Thus, we will consider sequences of information structures whose supports approach zero, with distributions of the form

$$\overline{\rho}(y) = \begin{cases} Ae^y y^{-h/g} & \text{if } \epsilon \leqslant y \leqslant \overline{y}; \\ 0 & \text{otherwise,} \end{cases}$$

where $\overline{y}$ will be determined later, ϵ is small, and A is a normalizing constant.

Now, it would seem that the potential of this information structure 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 and does not have a derivative at the boundary $x_1 + x_2 = \overline{y}$. This example therefore illustrates that one cannot always rely on differentiability to understand the continuum limit of the discrete IVO.

To work out $\gamma(x, (q, t))$ when $x_1 + x_2 = \overline{y}$, we can appeal to a discrete approximation. When k is large, the probability mass assigned to each of the boundary signal profiles with $x_1 + x_2 \leqslant \overline{y} \leqslant x_1 + x_2 + 1/k$ is approximately $\overline{\rho}(\overline{y})/k^2$. Recalling that the Lagrange multiplier is k, the IVO at each boundary signal profile is

$$\left(\overline{\rho}(\overline{y})/k^2 - k(0 - \overline{\rho}(\overline{y})/k^2)\right) gq,$$

which is maximized with $q = 1$. In addition, when ϵ is small, the number of boundary signal profiles is approximately $k\overline{y}$. Plugging in the normalizing constant in $\overline{\rho}$, we conclude that the expected highest IVO is approximately

$$\frac{\overline{y}e^{\overline{y}}\overline{y}^{-h/g}g}{\int_{y=\epsilon}^{\overline{y}}(y-\epsilon)e^y y^{-h/g}dy}. \tag{20}$$

If $h/g < 1/2$, then as $\epsilon \to 0$, this expression converges to

$$\frac{e^{\overline{y}}\overline{y}^{1-h/g}g}{\int_{y=0}^{\overline{y}} e^y y^{1-h/g}dy}. \tag{21}$$

We can further determine $\overline{y}$ by minimizing (21), although the correct value for $\overline{y}$ also follows from our subsequent analysis of the guarantee maximizer and complementary slackness.

By contrast, if $h/g \geqslant 2$, then the expected highest IVO (20) converges to zero. Thus, when ϵ is small, there is practically no trade in any equilibrium of any budget-balanced mechanism, even though there is a common knowledge of gains from trade.

For comparison, consider Akerlof (1970)'s lemons information structure, where the seller knows the state and the buyer has no information. In that model, trade always takes place in the low state, but it will not take place in the high state if the expected buyer value less than the seller's high value, i.e., $(h+g)/2 + g/2 < h$. But this is the same condition $h/g > 1/2$ under which the min potential is zero. Thus, efficient trade can be much harder to achieve than suggested by the lemons model.

3.6.2 Guarantee-maximizing mechanism

When $h/g \geqslant 2$, the min potential is zero, which every mechanism trivially guarantees. For $h/g < 2$, it takes more work to construct a guarantee maximizer. Again, we will maximize the expected lowest SVO across mechanisms of the form $(\mathbb{R}^2_+, q, t)$, where $q : \mathbb{R}^2_+ \to [0,1]$ and $t : \mathbb{R}^2_+ \to [-t_{max}, t_{max}]$ are differentiable. The associated SVO is:

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

To motivate the form of the guarantee maximizer, we can again appeal to complementary slackness. Recall that $\nu(x)$ was interior and $\overline{\rho}$ was strictly positive at almost all x for which $x_1 + x_2 \leqslant \overline{y}$ (where the boundary $\overline{y}$ is still to be determined). Hence, by condition (16), it should be that for every θ, x minimizes the SVO if and only if $x_1 + x_2 \leqslant \overline{y}$, i.e., there are constants L_θ such that $\lambda(\theta, x) = L_\theta$ when $x_1 + x_2 \leqslant \overline{y}$, and $\lambda(\theta, x) \geqslant L_\theta$ when $x_1 + x_2 > \overline{y}$. Moreover, since $\gamma(x, (q,t))$ is uniquely maximized by $q = 1$ when $x_1 + x_2 = \overline{y}$, by condition (15), it should be that $q(x) = 1$ when $x_1 + x_2 = \overline{y}$.

Numerical solutions of program (10) suggest a trade probability of the form $q(x) = q(x_1 + x_2)$, where q is increasing and $q(0) = 0$, and a transfer rule

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

A mechanism of this form is participation secure: with $x_1 = 0$ the seller can guarantee that any trade occurs at a high price of h, and with $x_2 = 0$ the buyer can guarantee that any trade occurs at a low price of g. Brooks and Du (2023) refer to this as a *proportional-price trading mechanism*. In such a mechanism, agents face a natural trade-off: higher x_i is associated with higher probability of trade but on worse terms for agent i.

Substituting these functional forms into the SVO, we get

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

Thus, $\lambda(\theta, x)$ is actually independent of θ, so we can simplify notation and define a single constant $L > 0$ so that $\lambda(\theta, x) = L$ for all θ and $x_1 + x_2 < \overline{y}$. Subject to the boundary condition $q(0) = 0$, this gives us a first-order ODE for q:

$$g(q(y) + q'(y)) - \frac{(h-g)q(y)}{y} = L.$$

The solution is of the form

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

Now, to satisfy complementary slackness, the allocation also has to hit 1 at $\overline{y}$ (which we can always achieve by scaling L up or down). Above that threshold, we will set $q = 1$, so that $q' = 0$, and hence the SVO changes to $g - \frac{h-g}{y}$. For aggregate actions above $\overline{y}$ to not induce a strictly lower SVO, it is necessary that $g - (h-g)/\overline{y} \geqslant L$. But just below the boundary, the SVO is exactly L, and $q(y) \approx 1$ and $q'(y) \geqslant 0$. Thus, it must be that $q'(y) \to 0$ as $y \to \overline{y}$, so that in fact $L = g - (h-g)/\overline{y}$.

The equation $q(\overline{y}) = 1$ implicitly defines $\overline{y}$ as the solution to

$$1 = \frac{g - \frac{h-g}{\overline{y}}}{g} e^{-\overline{y}} \overline{y}^{(h-g)/g} \int_{z=0}^{\overline{y}} e^{z} z^{-(h-g)/g} dz. \tag{22}$$

With the value of $\overline{y}$ pinned down, we have also completed the specification of the signal distribution, the allocation rule, and the transfer rule according to the formulae given above. We denote the resulting information structure by $\overline{I}_\epsilon$ and the resulting mechanism by $\overline{M}$.

Rearranging (22), the expected lowest SVO of $\overline{M}$ is

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

which is the same expression as (21) for the limit expected highest IVO of $\overline{I}_\epsilon$ as $\epsilon \to 0$. Hence, the expected lowest SVO and the limit expected lowest IVO coincide, so that again $P^* = G^*$, and we have constructed a guarantee maximizer and approximate potential minimizers.

3.7 On the Choice of Units

In these two examples, we solved for guarantee maximizers, which we deduced must have very particular functional forms. Now, suppose we take a guarantee maximizer M and use it to define a new mechanism M', where each agent takes an action $a_i \in \mathbb{R}_+$, and for some strictly increasing continuous functions $f_i : \mathbb{R}_+ \to \mathbb{R}_+$, the outcome is that of M under the action profile $(f_1(a_1), f_2(a_2), \ldots, f_N(a_N))$. This is clearly a superficial change of units, and so M' also maximizes the guarantee. Then why did the theory seem to select M?

The answer goes all the way back to dual reductions. When we reduced a mechanism, some of the Lagrange multipliers on obedience constraints were given to us, but we chose the multiplier on the "null" constraint—deviating from an action to itself—so that multipliers would sum to the same constant, regardless of the equilibrium action. And we used this constant as a Lagrange multiplier in certifying that the dual reduction had a higher guarantee. This normalization was carried through the analysis of the bounding programs, and even as we took the multiplier to infinity, where the labels for actions and the multiplier were balanced just so that in the continuum limit, the multiplier on the differential form of the SVO had a Lagrange multiplier of one.

This may seem strange that we fixed the multipliers and then optimized the mechanism. Should not the optimal multipliers depend on the mechanism? Certainly that is true if we are considering the multipliers that pin down the guarantee for an arbitrary mechanism. But we only need the multipliers to be optimal for the guarantee maximizer, and the theorem shows that there is a guarantee maximizer for which these normalized multipliers are optimal.

We could of course have chosen a different normalization of local obedience multipliers, and that would have selected different guarantee maximizers with a different choice of units. The same is true of the potential minimizer. While the unitary multipliers have selected appealing functional forms in these examples, it may be that in other applications a different normalization will lead to more intuitive or natural solutions.

3.8 The Literature

We conclude this section with a brief discussion of the literature. The potential of an information structure is simply the value of the standard Bayesian mechanism design problem, for which a seminal work is Myerson (1981) on optimal auctions with private values. Bergemann, Brooks, and Morris (2017) computed the revenue guarantee of the first-price auction in a setting with interdependent values. For auctions with pure common values, Du (2018) constructed mechanisms that extract approximately all of the surplus as the number of agents goes to infinity. The argument uses a version of the lower bounding program $\underline{G}(X(k^2))$. Subsequently, Bergemann, Brooks, and Morris (2016) used both bounding programs to construct a saddle point for the common value auction with two agents and binary values. Brooks and Du (2021b) solved the general problem with many agents and any value distribution.

Brooks and Du (2024) generalized the bounding programs to other mechanism design problems, and proved that there is no duality gap for a class of auction problems. Brooks and Du (2025) used dual reduction to provide a foundation for the bounding programs.[22]

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

[22] 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.

all of the surplus as the number of agents grows large. Brooks and Du (2023) constructed saddle points for a public expenditure problem, which in the case of two agents is equivalent to the bilateral trade problem described above.

4 Other Approaches to Robust Mechanism Design

We now shift gears and relate guarantee maximization to the rest of the literature on robust mechanism design. For recent surveys of this literature, see Bergemann and Morris (2012) and Carroll (2019). Our discussion focuses on those lines of research that are closest to the subject of this article.

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 an exact description of the environment. Moreover, we typically suppose that agents play the equilibrium that the designer most prefers. The known information structure, favorable equilibrium selection, as well as the assumption that the designer can implement a rich set of mechanisms, together imply the *revelation principle*: the designer can without loss restrict attention to mechanisms in which each agent's action is a report of their private information, and in equilibrium, agents report truthfully (Myerson, 1981, 1986). By and large, it is also typically assumed that the information structure satisfies the common prior assumption.

This model has been critiqued 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 the knife-edge case of 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 classic 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 higher-order beliefs, either on the part of the designer or the agents, is to use mechanisms that have equilibria in dominant strategies when values are private (i.e., each agent knows their ex post preferences) or in ex post equilibrium when there is interdependence. Dominant strategy and ex post implementation have been widely adopted in response to the Wilson critique.

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' *payoff types*, which parametrize all agents' and the designer's ex post preferences, and *belief types* that determine higher-order beliefs but are payoff irrelevant. They give a "separability" condition on preferences of the agents and the social choice correspondence under which the correspondence can be implemented for all belief types if and only if it can be implemented ex post with respect to the payoff types. Bergemann and Morris (2005) also give examples where separability fails and social choice correspondence is implementable, but not with ex post incentives. Jehiel et al. (2006) and subsequent papers have argued that, generically, the only social choice correspondences that can be implemented ex post are those that are constant.

Ex post implementation does not rely on common knowledge of higher-order beliefs, but it does depend on common knowledge of the payoff types and how they parametrize preferences. In the case of Dasgupta, Hammond, and Maskin (1979), these payoff types are the possible ex post preferences of the agents. In Bergemann and Morris (2005), the relationship between the payoff types and ex post preferences may be more complex.

In a somewhat different take, Chung and Ely (2007) studied auctions with correlated private values, where the seller's belief is fixed. The seller evaluates a given mechanism by the worst-case across all information structures in which agents know their values, and where the agents play the seller-preferred equilibrium. For "regular" value distributions, Chung and Ely construct an information structure in which the seller can do no better than the best dominant strategy mechanism.[23] 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 to other mechanism design problems. These papers also give examples where regularity fails, and the designer can uniformly improve on the best dominant strategy mechanism. There are also examples where the designer can do strictly better if the agents' beliefs satisfy the common prior assumption.

Like Chung and Ely (2007), guarantee maximization presumes the designer has expected utility preferences over states and outcomes, and participation constraints also play a central role in the theory. But in contrast to the literature on ex post implementation, guarantee maximization assumes 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, 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 are the worst-case are themselves plausible.

Most importantly, guarantee maximization does not make any primitive assumptions about the structure of the agents' private information. In particular, it is possible that agents know nothing at all about the state of the world, including their own preferences, except for the prior distribution. Hence, there is no private information or payoff types

[23]The substantive implication of Chung and Ely's 2007 regularity assumption is that the only binding constraints at the optimal dominant strategy mechanism are local downward incentive compatibility and interim participation for the lowest value.

with respect to which the designer could provide incentives ex post. While this makes the designer's problem harder, it also makes the guarantee more robust.

4.2 Equilibrium Selection and Strong Nash Implementation

As we have said, the dominant paradigm in Bayesian mechanism design is to select the equilibrium that is most preferred by the designer. Without this assumption, the revelation principle would not apply, and we would have to grapple with optimization over indirect mechanisms and their equilibria, a decidedly daunting analytical task.

In contrast, the literature on mechanism design under complete information has predominantly asked for desirable outcomes in *all* equilibria (Maskin, 1999). This is so-called *strong* (or *full*) *implementation*, to be contrasted with *weak* (or *partial*) *implementation* in which the designer picks the equilibrium. Strong implementation has also been studied in a Bayesian setting by Serrano and Vohra (2010) and others. The mechanisms that achieve strong implementation often feature devices such as integer games 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.

Guarantee maximization 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.[24]

Is equilibrium multiplicity a practical concern or a theoretical nuisance? In the second-price auction with private values, each agent has a unique weakly undominated strategy in which they bid their value (Vickrey, 1961). But there are also "bidding ring" equilibria in which one agent makes a very high bid, and the others essentially refuse to participate. By contrast, for independent private values, the payoff-equivalent equilibrium of the first-price auction is essentially unique (Lizzeri and Persico, 2000). Rothkopf, Teisberg, and Kahn (1990) have argued that the bidding ring equilibria are a realistic method of collusion in repeated auctions, wherein the agents take turns as the high bidder, and that as a result, second-price auctions are more vulnerable to collusion than first-price auctions. At the same time, the designer could always perturb the second-price auction by adding a noisy hidden reserve price, so that bidding one's value becomes strictly dominant. Moreover, this can be done with arbitrarily small probability, and at negligible cost to the seller. Would such perturbations actually deter collusive behavior? The answer no doubt depends 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 under the idealized and standard assumption that agents are sensitive to arbitrarily small changes in their payoffs, because a mechanism that weakly implements the desired outcome can always be perturbed at negligible cost in a manner that eliminates undesirable equilibria. At the same time, there may be more than one way of selecting favorable equilibria. In a similar spirit, 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.

[24] Glazer and Rosenthal (1992) argue that due to the complex structure of Abreu-Matsushima mechanisms, real human agents would not play them in a manner consistent with the theory. See Abreu and Matsushima (1992a) for a response. Kapon, Del Carpio, and Chassang (2024) view a variant of the Abreu-Matsushima targeted punishments as a practical tool for mechanism design in large populations.

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 common priors and the hypothesis that agents will play local best responses. In that sense, the optimal guarantee does not rely on the full force of equilibrium.

4.3 Robustness to Fundamentals

In defining the guarantee in Section 2, we fixed the marginal distribution over the payoff-relevant state θ, so that the "robustness" in the guarantee is only with respect to the agents' information and equilibrium, and not with respect to fundamentals. There is a large body of work that instead emphasizes robustness with respect to fundamentals, and maintains classical assumptions on information and equilibrium selection. For example, Carroll (2017), Che and Zhong (2021), and Deb and Roesler (2023) study multidimensional screening, assuming that the agent knows their private values but with uncertainty in the correlation in values across dimensions. In multi-agent settings, He and Li (2022) analyzes guarantees of second-price auctions with (possibly random) reserve prices where agents know only their own values but the joint distribution of values is unknown to the designer. Similar exercises are conducted by Che (2020) for a class of "competitive" auctions and by Zhang (2022) for maximizing fees in bilateral trade.

These papers analyze either single agent problems with general mechanisms, or multi-agent problems with private values and dominant strategy mechanisms. In either case, higher-order beliefs are irrelevant to the analysis. Guarantee maximization can be applied in single-agent models, and a promising avenue is to apply the theory to multi-product monopoly. But in our view, 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 on algorithmic mechanism design in which 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 the distribution of fundamentals. See Nisan et al. (2007) for a textbook treatment.

Of course, one could ask for robustness with respect to both fundamentals and information. Brooks (2013) conducts such an analysis of private value auctions where only certain moments of the value distribution are known. Within the framework described in this article, it is straightforward to incorporate Knightian uncertainty about payoff-relevant states. Indeed, Brooks and Du (2021b, 2024) discuss how the performance of a guarantee maximizer would vary with the prior, and prove that there is a lower bound on the mechanism's guarantee that is linear in the true prior. Also, as discussed in Brooks and Du (2024, 2023), the results on guarantees in bilateral trade that we reviewed in Section 3 remain valid even if the designer only knows a lower bound on the gains from trade. Similarly, Brooks, Du, and Feffer (2025) consider optimal revenue guarantees in auctions when the designer only knows the expectation of each agent's value.

For mechanism design to be useful as a normative theory, the parameters of the model must be quantities that we could reasonably expect a practitioner to be able to specify, either through empirics or introspection. The worst-case criterion can be a useful modeling device for "solving out" those parameters that a practitioner is unable to quantify but on which the theory depends. For that reason, it may be advantageous to take as the primitive a set of priors that reflect the designer's best understanding of the environment, while also allowing for Knightian uncertainty about those aspects of fundamentals of which the designer is uninformed.

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 vulnerabilities and deficiencies. The hope is that such critical analysis will lead to a more robust solution to the problem of mechanism design. After having considered the potential benefits of this approach, it seems only fitting that we should similarly turn a critical eye on the theory itself, and assess the ways in which guarantee maximization 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 that evaluating a mechanism by its the worst-case outcome is too extreme, and the worst case may not be the one that is practically relevant. For example, suppose that the guarantee is maximized by a mechanism that achieves a payoff of 1 in all information structures and equilibria. Another mechanism could achieve a payoff of 2, except at a single information structure and equilibrium, in which it achieves a payoff of 0. Guarantee maximization selects the first mechanism, but the second mechanism might be preferred, if the bad information structure/equilibrium is a poor reflection of the environments in which the mechanism is expected to operate.

We do not take guarantee maximization literally as the preferences of the mechanism designer, and we do not think that a mechanism is good just because it maximizes the guarantee. Rather, it is a heuristic that we hope will guide us to novel and efficacious solutions. And we have to look at the particular mechanisms that arise from the theory, and assess whether the incentive structures driving their performance would be effective with actual human beings as the agents, and not just their theoretical idealizations. We should also look at those worst-case environments that the mechanisms are implicitly guarding against; if those environments seem plausible, then that is a further argument that the guarantee-maximizer is a prudent design.

For example, proportional auctions distribute 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 feature if we are worried that information might facilitate coordination on who is taking which action in a way that could lower revenue, such as by loading the allocation on the agent with the most optimistic signal, thereby generating a strong winner's curse that depresses willingness to pay (as can happen in a first-price auction). Moreover, proportional auctions 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. This model is stylized, but not in a manner that seems to do any more violence to reality than other information structures that are widely used in the literature for their balance of realism and tractability.

But suppose it turns out that the potential-minimizing information structure is one that seems unrealistic. One way to proceed would be to restrict the set of information structures to exclude the potential minimizer and any other information structures that share its objectionable properties. For example, in common value auctions, the potential-minimizing information structure can have 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 of such bounds 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. This is especially relevant for problems in which there are no participation constraints, such as taxation or voting. Without participation constraints, the min potential 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 it seems reasonable that the designer may wish to rule out this degenerate case, and only consider information structures in which the agents have a non-negligible amount of information.

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 assumed to know their values, and we suggested that it can be thought of as a worst case of sorts. As we have seen, potential minimizers in optimal auctions will have independent signals, but they generally do not exhibit private values. One view is that private values is a technically convenient but unnatural assumption, and we should develop alternative theories of optimal auctions that do not rely on private values. However, one could also argue that agents know more about themselves than they do about others, and private values express this idea in its most extreme form.

In fact, Bergemann and Morris (2016) incorporate 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, but they may observe more. The theory of optimal guarantees 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 such a baseline information structure. They show that 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).

In optimal auctions, Brooks and Du (2025) show that when values are independent (an assumption on the prior μ), the independent private value model does indeed minimize the potential among all private value information structures (where again, agents know their own values but may know more). However, if the values are correlated, then the information structure in which the agents only learn their value does not minimize the potential, and other information structures will lead to lower revenue in all mechanisms and equilibria. A similar point has been recently made by Pusztai and Rahman (2025). These results provide an informational robustness foundation for the independent private value model of Vickrey (1961) and Myerson (1981).

One could of course combine lower bounds and upper bounds, and this possibility is also explored in Brooks, Du, and Haberman (2024a), by placing bounds on how much information an agent might have about the state and others' types, beyond their base type, where information is quantified with a divergence. This effectively allows us to explore the possible equilibrium outcomes in a "ball" around the baseline type space, where we still allow arbitrary forms of correlation in actions that do not affect the beliefs of the agents (or of subsets of agents). This is dual to the approach taken in Hansen and Sargent (2001) and Maccheroni, Marinacci, and Rustichini (2006), where the designer penalizes models by how far they are from a given benchmark, e.g., in terms of relative entropy.

5.2 Participation Security is Too Strong

It is clear that participation constraints play a central role in the theory. But there are different approaches one can take to modeling participation. In this article, we have described two: participation security of mechanisms and interim individual rationality for information structures and equilibria. 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 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 each agent, which is either 0 or 3/4, with likelihoods ϵ and $1-\epsilon$ respectively. In this setting, if an agent takes a secure action, their contribution must be zero. As a result, if only one agent's action is not secure, they must cover the whole cost of the public good. Any participation-secure mechanism therefore 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 efficient surplus of $3/2-1=1/2$, uniformly across information structures.[25]

Thus, participation security is in general more demanding than the requirement that for every information structure there exists an equilibrium with non-negative interim utility. Further advances in settings like this public good example 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? In our baseline model, there are finitely many outcomes, and conditional on the action profile, the mechanism produces a lottery over those outcomes. Both proportional auctions and proportional-price trading mechanisms generate lotteries that respond smoothly to the agents' actions. This structure is not entirely surprising. Guarantee maximizers tend to equalize the strategic virtual objective across action profiles, and doing so requires interior solutions (similarly to how maxmin mixed strategies in zero-sum games may use randomization to equalize an opponent's payoffs across their actions). One may be skeptical of a designer's ability to commit to such randomization, especially with such carefully calibrated

[25]The worst-case information structure for this mechanism would be one for which the interim expected value 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 $\epsilon \to 0$.

probabilities. However, if the goods being allocated are divisible, e.g., bushels of wheat or shares of common stock, then we can reinterpret the lottery as allocating a portion of the good to each agent, which seems easier to implement in practice.

One might ask, in addition to randomization within the mechanism, could the designer achieve an even higher guarantee randomizing over the mechanism itself?[26] This seems to have been ruled out by assumption in Section 2. However, randomization over mechanisms is always *equivalent* to randomization within the mechanism: Instead a lottery over mechanisms, the designer could put probability one on the lottery's strategic normal form, in which each agent's pure behavioral strategies specify an action for each outcome of the lottery. A similar comment applies to information: A lottery over information structures is equivalent to a single information structure, with different common knowledge components corresponding to different outcomes of the lottery.

If randomization within a mechanism is seen as problematic, then one way to proceed is to limit the domain of mechanisms over which the guarantee is maximized. This is the approach taken by Bergemann, Brooks, and Morris (2019), who optimized revenue guarantees 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 information, Bergemann, Brooks, and Morris (2019) also consider a restricted guarantee where the information must be a symmetric affiliated values environment and the agents play the monotone pure strategy equilibrium of Milgrom and Weber (1982). The guarantee of the first-price auction is unchanged (since its worst-case is in this class), but now the second-price auction's guarantee is the same as that of the first-price auction. The potential minimizer is not regular and in fact non-local equilibrium constraints bind. An interesting direction for future research is to explore the relationship between randomization

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

and other patterns of binding equilibrium constraints, which correspond to different notions of the strategic virtual objective.

5.4 Simplicity and Portability?

We motivated guarantee maximization 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 that the theoretical incentives will be effective in shaping the behavior of real human beings. We argued that the search for the worst-case information would lead us to an ordered structure on signals that reflects the most potent form of information to amplify the threat of manipulation and objection. Such structure is also reflected in the guarantee-maximizing mechanisms, and the theoretical argument underlying the optimal guarantee only relies on the fact that agents' play local best responses. Moreover, when a saddle point exists, guarantee maximizers will always be as or more effective in achieving the designer's goals as they are in the potential minimizers that rationalize them as optimal. In that sense, guarantee-maximizing mechanisms exhibit a strong form of portability across environments.

One can push back against these claims. The local structure of equilibrium constraints is an aspect of simplicity, but it is not every kind of simplicity that we might care about. It also matters whether the functional form of the mechanism is easy to interpret and explain to users. In the solved examples, the functional forms seem relatively simple and low dimensional, but this need not always be the case. Regarding portability, we have already remarked that the guarantee is a rather extreme notion. What really matters is portability across environments that are empirically relevant, and even then, portability should be traded off against optimality in environments that are regarded as more likely to arise.

Perhaps a more serious issue is that we have failed to articulate precisely when and why incentive structures that are effective in theory may not be effective in practice. We have appealed to complexity as a central issue. But complexity is an aspect of not just the mechanism but also the information and equilibrium. Even if the mechanism is simple, we should still be concerned that agents may not arrive at common knowledge or best responses if other aspects of the environment are too complex, and this could ultimately undermine the validity of the guarantee. It is true 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. In that sense, the guarantee does not rely on the full force of equilibrium. Still, it seems we have avoided a central issue, which is exactly what kinds of complexity will make it hard for agents to learn how to play a mechanism well and converge to equilibrium.[27]

Guarantee maximization is a workaround for not having been able to properly define or quantify simplicity or portability, so that these desiderata could be directly incorporated into the design problem. Echoing one of our earlier sentiments, the best hope for this theory is that in spite of not having fully described what we are after, it may lead us there anyway, and we will it when we see it. At the very least, we hope that the approach will generate novel ideas for how to structure incentives. It may be, though, that the ultimate progress of mechanism design depends on more directly addressing the issue of how agents reason through a mechanism and converge to strategies, so that we may design institutions with greater regard for the realities and peculiarities of human strategic behavior.

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.

[27]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.

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.

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

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

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.

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.

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.

Hansen, L. and T. J. Sargent (2001): "Robust control and model uncertainty," *American Economic Review*, 91, 60–66.

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

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.

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.

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

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity aversion, robustness, and the variational representation of preferences," *Econometrica*, 74, 1447–1498.

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.

Myerson, R. B. and M. A. Satterthwaite (1983): "Efficient mechanisms for bilateral trading," *Journal of economic theory*, 29, 265–281.

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.

Pusztai, P. and D. M. Rahman (2025): "Redistributing Information Rents," *Available at SSRN 5160344*.

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*.

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