8-2024

# A robust optimization approach to mechanism design

Jiangtao LI
*Singapore Management University*, jtli@smu.edu.sg

Kexin WANG
*Singapore Management University*, kexinwang.2020@phdecons.smu.edu.sg

# A Robust Optimization Approach to Mechanism Design*

Jiangtao Li† Kexin Wang‡

August 12, 2024

**Abstract**

We study the design of mechanisms when the mechanism designer faces local uncertainty about agents' beliefs. Specifically, we consider a designer who does not know the exact beliefs of the agents but is confident that her estimate is within $\epsilon$ of the beliefs held by the agents (where $\epsilon$ reflects the degree of local uncertainty). Adopting the robust optimization approach, we design mechanisms that incentivize agents to truthfully report their payoff-relevant information regardless of their actual beliefs. For any fixed $\epsilon$, we identify necessary and sufficient conditions under which requiring this sense of robustness is without loss of revenue for the designer. By analyzing the limiting case in which $\epsilon$ approaches 0, we provide two rationales for the widely studied Bayesian mechanism design framework.

*We are grateful to Tilman Börgers, Yi-Chun Chen, Ying-Ju Chen, Kim-Sau Chung, Gregorio Curello, Huiyi Guo, Wei He, Yunan Li, Anton Kolotilin, Stephen Morris, Mariann Ollár, Antonio Penta, Hien Pham, Ludvig Sinander, Qinggong Wu, Takuro Yamashita, and participants at various seminars and conferences for helpful comments.

†School of Economics, Singapore Management University, jtli@smu.edu.sg

‡School of Economics, Singapore Management University, kexinwang.2020@phdecons.smu.edu.sg

# 1 Introduction

Many classical analyses of mechanism design problems adopt the Bayesian approach, typically assuming that agents' belief about other agents are common knowledge among the agents and the designer. Under this assumption and using the solution concept of Bayesian Nash equilibrium, the optimal Bayesian mechanism design problem can be solved by maximizing the designer's expected payoff subject to the Bayesian incentive constraints—each agent finds it optimal to truthfully report her payoff-relevant information in expectation. This assumption, while standard in Bayesian mechanism design models, is nevertheless very strong. Consequently, the theoretical conclusions can sometimes be fragile; mechanisms optimized to perform well when this assumption is exactly true may still fail miserably in the much more frequent cases when this assumption is untrue. The so-called Wilson doctrine suggests that practical mechanisms should be designed without assuming detailed knowledge of the underlying economic environment. Relaxing these assumptions has been the focus on the recent literature on robust mechanism design.

In response to Wilson's critique, another common approach uses the stronger solution concept of dominant strategy Nash equilibrium. This approach does not rely on any information about the agents' beliefs, and the optimal dominant strategy mechanism design problem can be solved by maximizing the designer's expected payoff subject to the dominant strategy incentive constraints—each agent has a dominant strategy to truthfully report her payoff-relevant information. While such an approach minimizes the impact of any assumption about agents' beliefs, for many mechanism design problems, the class of dominant strategy mechanisms is quite small and only includes mechanisms that are rather unattractive for the designer. This comes as no surprise, since the designer is not allowed to utilize any information about the agents' beliefs in designing mechanisms.

In many realistic settings, the designer may have some (imprecise) understanding of agents' beliefs. Although the designer does not know the agents' exact beliefs, she could nevertheless form an estimate of the agents' beliefs, perhaps based on historical data from similar interactions in the past. The designer is aware that her estimate of agents' beliefs could be wrong, but believes that her estimate is close to the true beliefs of the agents. This paper studies a robust optimization approach to mechanism design for such settings.

We consider a general social choice environment with quasi-linear preferences and independent private values. The designer has a benchmark model represented by a distribution $\mu$ of the agents' payoff-relevant information and a benchmark belief $\mu_i$ for each agent $i$. The designer faces local uncertainty about the agents' beliefs—while the designer does not know the agents' exact beliefs, she is confident that the agents' beliefs are $\epsilon$-close to those derived from the benchmark prior $\mu$ (where $\epsilon$ reflects the degree of local uncertainty in the estimate). In other words, rather than being fixated on a specific belief $\mu_i$ for each agent $i$, the designer in our model entertains a range of possible beliefs that are $\epsilon$-close to $\mu_i$. In contrast to the standard Bayesian approach which only requires each agent to truthfully report her payoff-relevant information in expectation with respect to a single belief, the designer seeks to design a mechanism in which each agent has an incentive to truthfully report her payoff-relevant information in expectation regardless of the actual belief held by each agent, as long as the belief is $\epsilon$-close to that derived from the benchmark prior.

The robust optimization approach has several appealing features. In comparison to the Bayesian approach, it is robust to slight misspecification about agents' beliefs. In comparison to the dominant strategy mechanism design approach, it utilizes information about the agents' beliefs and thus could potentially achieve a higher expected revenue for the designer. To fix ideas and also to illustrate some of the motivations of our analysis, consider the following example.

**Example 1.** Consider the problem of designing a trading platform for two traders, $A$ and $B$, with the goal of maximizing intermediation profit. Each trader can buy or (short) sell one unit of the asset and has private information about her valuation for the good.[1] The platform cannot hold inventory (ex post market-clearing is imposed). The set of possible types for trader $A$ is $\Theta_A = \{0, \frac{2}{3}\}$, and the set of possible types for trader $B$ is $\Theta_B = \{\frac{1}{3}, 1\}$. Each agent may be either the buyer or the seller, depending on the realization of the privately observed information and the choice of the mechanism: the agent's role as the buyer or the seller is endogenously determined by his report and cannot be identified prior to trade. The platform has the following estimate of the distribution $\mu$ of the traders' types:

[1]For related models, see Cramton, Gibbons, and Klemperer (1987), Lu and Robert (2001), Chen and Li (2018), Loertscher and Marx (2020), and Li and Dworczak (2024).

| | $\theta_B^1 = \frac{1}{3}$ | $\theta_B^2 = 1$ |
|---|---|---|
| $\theta_A^1 = 0$ | $\frac{1}{12}$ | $\frac{1}{24}$ |
| $\theta_A^2 = \frac{2}{3}$ | $\frac{7}{12}$ | $\frac{7}{24}$ |

**The standard Bayesian approach.** Assuming that the joint distribution of the traders' valuation is common knowledge among the traders and the platform, the designer chooses the optimal Bayesian incentive compatible mechanism, which generates an expected revenue of $S^B = 0.361$. However, a mechanism that satisfies Bayesian incentive compatibility with respect to the benchmark prior $\mu$ is not robust to misspecification of agents' beliefs. When the agents' actual beliefs are not consistent with the benchmark prior $\mu$, truth-telling might not constitute a Bayesian Nash equilibrium in the mechanism.

**The dominant strategy mechanism design approach.** To eliminate the reliance on the assumption about agents' belief, the platform might use a dominant strategy mechanism. The optimal dominant strategy mechanism generates an expected revenue of $S^D = 0.306$. While the optimal dominant strategy mechanism does not reply on any assumptions about agents' beliefs, the designer has to incur loss in expected revenue, that is, $S^D < S^B$.

**The robust optimization approach.** Let $\mu_i(\cdot)$ denote the benchmark belief of agent $i$ derived from the benchmark prior $\mu$. The designer is aware that her estimate of the agents' beliefs (as derived from the benchmark prior) could be different from the actual beliefs of the agents, but is confident that her estimate is close to the true beliefs. For each agent $i$, let

$$U_i^\epsilon := \left\{ \nu_i \in \Delta\Theta_{-i} : \max_{\theta_{-i} \in \Theta_{-i}} \left| \nu_i(\theta_{-i}) - \mu_i(\theta_{-i}) \right| \leq \epsilon \right\}$$

denote the collection of beliefs that are perceived to be plausible to the designer. As long as $\epsilon \leq \frac{1}{6}$, then the optimal robust incentive compatible mechanism achieve the expected revenue of $S^R(\epsilon) = 0.361$. Thus, requiring robustness to local uncertainty about agents' beliefs (for a range of $\epsilon$ values) is without loss of revenue for the designer.

For any fixed $\epsilon$, we identify necessary and sufficient conditions under which requiring this sense of robustness is without loss of revenue for the designer. This result builds on the recent literature on the network approach to mechanism design and specifically the dual interpretation of a revenue maximization problem as a network flow problem; see for example Rochet (1987), Heydenreich, Müller, Uetz, and Vohra (2009), and Vohra (2011). Given a decision rule $q$ and a possible belief $\nu_i$ of agent $i$, we define a network $G(q, \nu_i)$ where the set of nodes is the set of types for agent $i$ plus a dummy type (which corresponds to opting out of the mechanism), and the length of a directed edge between two types is defined based on the corresponding incentive constraint. The revenue maximization problem has a dual interpretation of a network flow problem, where the optimization problem reduces to determining the shortest-path tree (the union of shortest paths from the dummy type to all nodes) for each agent. We say that a decision rule $q$ satisfies the uniform shortest-path tree property with respect the uncertainty set if for each agent $i$, there is the same shortest-path tree for all networks $G(q, \nu_i)$ where $\nu_i$ is a possible belief of agent $i$. Theorem 1 shows that requiring robustness to local uncertainty about agents' beliefs is without loss of revenue for the designer if and only if there exists an optimal decision rule $q^*$ of the Bayesian mechanism design problem such that $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set.

The uniform shortest-path tree property with respect to an uncertainty set is of interest because a number of resource allocation problems satisfy this condition. First, we consider environments in which the optimal decision rule of the Bayesian mechanism design problem satisfies the uniform shortest-path tree property with respect to the uncertainty set (for any $\epsilon$). The leading example is the class of one-dimensional environments. In these environments, it follows from Theorem 1 that $S^B = S^R(\epsilon) = S^D$ for any $\epsilon$.[2] Second, we consider environments in which the optimal decision rule of the Bayesian mechanism problem satisfies the uniform shortest-path tree property with respect to an uncertainty set for smaller degrees of local uncertainty even if it is not the case for larger degrees of local uncertainty. In these environments, by Theorem 1, we have $S^B = S^R(\epsilon) > S^D$ for a range of $\epsilon$ values. We illustrate this feature using the bilateral trade problem with ex ante unidentified traders.

More importantly, we examine the robust optimization approach in the context where the degree of local uncertainty converges to zero. This analysis is particularly significant as it allows us to provide a rationale for the Bayesian mechanism design

[2] $S^B$ (resp. $S^D$) denotes the highest expected revenue from the optimal Bayesian mechanism (resp. the optimal dominant strategy mechanism). $S^R(\epsilon)$ denotes the highest expected revenue from the optimal robust incentive compatible mechanism with respect to the uncertainty set.

framework. The motivation behind this is as follows: Although the Bayesian approach relies on strong common knowledge assumptions, it remains prevalent in the mechanism design literature. A common informal justification for this approach is that, while the designer recognizes the Bayesian model as merely an approximation, its use is validated by a wealth of historical transaction data. Within the robust optimization framework, the degree of local uncertainty decreases as the number of historical transactions increases, ultimately approaching zero when such data is abundant.

We provide two foundations for the Bayesian approach. Theorem 2 shows that, under a condition on the uniqueness of the shortest path, there exists a threshold $\bar{\epsilon} > 0$ such that the uniform shortest-path tree property with respect to the uncertainty set is satisfied for any $\epsilon \leq \bar{\epsilon}$. It follows from Theorem 1 that $S^B = S^R(\epsilon)$ for any $\epsilon \leq \bar{\epsilon}$. Theorem 3 shows that, under a mild Slater condition, the revenue loss due to robustness is vanishingly small as $\epsilon$ approaches 0. We interpret these results as foundations of the Bayesian mechanism design framework.

The rest of this introduction discusses related literature. Section 2 presents the notations, concepts, and the model. Section 3 characterizes decision rules that are robustly implementable with respect to an uncertainty set. Section 4 studies robustness to local uncertainty for a fixed $\epsilon$. Section 5 studies robustness to local uncertainty in the limiting case in which $\epsilon$ approaches zero, offering justifications for the Bayesian mechanism design framework. Section 6 concludes the paper.

## 1.1 Related literature

First and foremost, this paper contributes to the robust optimization literature where real-world optimization problems are often modeled as uncertain optimization problems; see for example Ben-Tal and Nemirovski (2002), Beyer and Sendhoff (2007), and Bertsimas, Gupta, and Kallus (2018). We contribute to this literature by systematically analyzing this robust optimization approach to mechanism design.

Our paper studies the design of mechanisms when the designer has non-Bayesian uncertainty about agents' beliefs, thereby contributing to the growing literature on robust mechanism design. A large body of papers in this literature focus on settings in which the designer has no information whatsoever regarding agents' beliefs; see for example Bergemann and Morris (2005), Chung and Ely (2007), Chen and Li (2018), and Yamashita and Zhu (2022).[3] In contrast to these papers, our paper considers local uncertainty about agents' beliefs, where each agent's true belief is within $\epsilon$ of those derived from the designer's benchmark prior. This leads to distinct set of results from those in the literature. For instance, Chung and Ely (2007) and Chen and Li (2018) show that the designer should use a dominant strategy mechanism, whereas our analysis supports the use of robust incentive compatible mechanisms.

More recently, there is increasing interest in designing mechanisms when the designer has partial knowledge of agents' beliefs. Ollár and Penta (2017) and Ollár and Penta (2023) work with general belief restrictions (such as moments conditions on the distribution and identical but unknown distributions) and study how belief restrictions can be used to design transfer schemes to achieve full implementation of certain decision rules. The most closely related paper to ours is Lopomo, Rigotti, and Shannon (2020). Like us, they work with a generalization of the standard Bayesian mechanism design problem in which each agent is associated with a set of beliefs rather than a unique belief. In the environment with one-dimensional types, they demonstrate that under certain conditions, such as the fully overlapping beliefs condition, and with the additional assumption that gross utilities increase with types, robust incentive compatibility implies an ex post envelope condition. This means that for any ex post incentive compatible decision rule, differences in ex post utilities across types are pinned down by the allocation rule. Therefore, they argue that robustness to arbitrarily small amounts of misspecification generates a discontinuity in the set of feasible mechanisms and uniquely selects simple, ex post incentive compatible mechanisms.

Our paper differs from Lopomo, Rigotti, and Shannon (2020) in several key aspects. First, we consider a specific form of uncertainty set that corresponds to local uncertainty about agents' beliefs. This type of uncertainty set is commonly used in the robust optimization literature, and the focus on this type of uncertainty allows us to derive more precise comparisons when comparing different classes of mechanisms. Second, while Lopomo, Rigotti, and Shannon (2020) examines the payment schemes necessary to implement an ex post incentive compatible decision rule given an uncertainty set, our focus is on revenue maximization problems. In one-dimensional environments, while their findings suggest that the payment schemes to implement any ex post incentive

[3] Also see Du (2018) and Brooks and Du (2021) that study the design of mechanisms when the designer does not know the information structure. In these papers, the designer also has minimal information about agents' beliefs, but the beliefs would have to be consistent with some information structure.

compatible decision rule are rather restrictive, our emphasis on revenue maximization enables us to demonstrate that requiring robustness to local uncertainty is actually without loss of revenue for the designer. Third, although our uncertainty set is a special case of theirs, we consider general social choice environments under multi-dimensional, independent and private types. In particular, multi-dimensional environments are not covered in their analysis. We show that in multi-dimensional environments, although the set of possible beliefs for each agent satisfies the fully overlapping condition, the set of feasible mechanisms that are robust incentive compatible can be much richer than the set of ex post incentive compatible mechanisms. Overall, while their findings suggest that the payment schemes to implement any ex post incentive compatible decision rule are restrictive within the environments they consider, our analysis underscores the advantages of the robust optimization approach in terms of the designer's revenue in a wide variety of environments. These two distinct perspectives are complementary and together offer a more comprehensive understanding of the robust optimization approach to mechanism design.

Another closely related paper is Cui, Chen, and Shen (2009). Like our work, their paper examines an auctioneer dealing with local uncertainty about agents' beliefs and explores robust incentive compatibility with respect to an uncertainty set. However, there are three key differences between the two studies. First, while their focus is on auctions, we work with a general social choice environment. Second, whereas they directly address decision rules that are robustly implementable with respect to the uncertainty set, we provide conditions under which decision rules can be robustly implemented. Third, a key part of our analysis is the examination of the limiting case where the degree of local uncertainty approaches zero, which we use to offer justifications for the Bayesian mechanism design framework. This aspect is not considered in their analysis.

Pham and Yamashita (2024) also address the issue of revenue maximization when the designer lacks full certainty about agents' beliefs. However, the key distinction lies in the modeling approach: while they focus on local uncertainty concerning the prior, our model centers on local uncertainty about the agents' interim beliefs. This difference in modeling leads to dramatically different results. Specifically, they observe that even with minimal local uncertainty in the prior, there is effectively no constraint on the agents' interim beliefs, leading them to advocate for dominant strategy mechanisms. The stark contrast between Pham and Yamashita (2024) and our paper highlights the crucial impact of the modeling choice.

Our paper also contributes to the strand of the mechanism design literature that shows the equivalence of different classes of mechanisms. Notably, Manelli and Vincent (2010) and Gershkov, Goeree, Kushnir, Moldovanu, and Shi (2013) establish the equivalence of Bayesian and dominant strategy mechanisms in social choice environments with linear utilities and independent, one-dimensional, private types. Their findings suggest that, among other things, the highest expected revenue from Bayesian mechanisms can be achieved via a dominant strategy mechanism. Our paper provides necessary and sufficient conditions under which the highest expected revenue from Bayesian mechanisms can be achieved via a robust incentive compatible mechanism (with respect to an uncertainty set).

There are several other papers that study local robustness in mechanism design settings, including Bergemann and Schlag (2011), Jehiel, Meyer-ter Vehn, and Moldovanu (2012), Carroll and Meng (2016), Madarász and Prat (2017), and Carroll (2017), among others.

# 2 Preliminaries

## 2.1 The benchmark model

There is a finite set $\mathcal{I} = \{1, 2, \ldots, I\}$ of risk-neutral agents and a finite set $\mathcal{K} = \{1, 2, \ldots, K\}$ of alternatives. Each agent $i$ has a type $\theta_i \in \mathbb{R}^K$ that represents her gross utility under the $K$ alternatives; $\theta_i(k)$ denotes agent $i$'s gross utility under the alternative $k$.[4] The set of possible types of agent $i$ is a finite set $\Theta_i \subset \mathbb{R}^K$. The set of possible type profiles is $\Theta = \prod_{i\in\mathcal{I}} \Theta_i$ with representative element $\theta$. We write $\theta_{-i}$ for a type profile of agents other than agent $i$, i.e., $\theta_{-i} \in \Theta_{-i} = \prod_{j\neq i} \Theta_j$. If $Y$ is a measurable space, then $\Delta Y$ is the set of all probability measures on $Y$. If $Y$ is a metric space, then we treat it as a measurable space with its Borel $\sigma$-algebra.

The designer has an estimate of the distribution $\mu \in \Delta\Theta$ of agents' types. We

[4] We may represent agents' types in different ways. For instance, when studying the single-unit auction or the bilateral trade with ex ante unidentified traders, it is more convenient to represent agent $i$'s type by $v_i \in \mathbb{R}$, denoting agent $i$'s value of the object.

refer to $\mu$ as the benchmark prior. For simplicity, assume that $\mu$ has full support. Let

$$\mu_i(\theta_i) = \sum_{\theta_{-i}\in\Theta_{-i}} \mu(\theta_i, \theta_{-i})$$

denote the marginal probability of $\theta_i$ and

$$\mu_i(\theta_{-i}|\theta_i) = \frac{\mu(\theta_i, \theta_{-i})}{\mu_i(\theta_i)}$$

the conditional probability of $\theta_{-i}$ given $\theta_i$ (derived from the benchmark prior $\mu$). We assume that types are independent under the distribution $\mu$, that is, $\mu(\theta) = \prod_{i\in\mathcal{I}} \mu_i(\theta_i)$ for any $\theta \in \Theta$. Since types are independent under the distribution $\mu$, for notational simplicity, we write $\mu_i(\theta_{-i}) = \mu_i(\theta_{-i}|\theta_i)$ for all $\theta_i$.[5]

## 2.2 Local uncertainty about agents' beliefs

The Bayesian approach assumes that agents' beliefs are derived from the benchmark prior and these beliefs are further assumed to be common knowledge among the agents and the designer. In contrast, the designer in our model has local uncertainty about agents' beliefs. The designer is aware that her estimate of the agents' beliefs (as derived from the benchmark prior) could be different from the actual beliefs of the agents, but is confident that her estimate is close to the true beliefs of the agents. We model a notion of uncertainty set that captures such local uncertainty about agents' beliefs.

For each agent $i$, let

$$U_i^\epsilon := \left\{\nu_i \in \Delta\Theta_{-i} : \max_{\theta_{-i}\in\Theta_{-i}} \left|\nu_i(\theta_{-i}) - \mu_i(\theta_{-i})\right| \leq \epsilon\right\}$$

denote the collection of all probability distributions on $\Theta_{-i}$ that are $\epsilon$-close to that derived from the benchmark prior. We interpret $U_i^\epsilon$ as the set of beliefs that are perceived to be plausible by the designer; the designer is confident that the agents' true beliefs are contained in the uncertainty set, but does not know or is unwilling to impose further assumptions about the true beliefs of the agents.

The parameter $\epsilon$ reflects the range of possible misspecification of the benchmark

[5] For correlated types, Lopomo, Rigotti, and Shannon (2022) study a robust version of the classic surplus extraction problem and show that the designer can achieve virtual extraction whenever agents' beliefs satisfy a natural set-valued analogue of the convex independence conditions in Cremer and McLean (1988).

model, or equivalently, the degree of local uncertainty about agents' beliefs. Clearly, the larger the value of $\epsilon$, the more ambiguity the designer has about agents' beliefs. When $\epsilon = 0$, the uncertainty set of each agent $i$ is just a singleton set—the conditional probability distribution derived from the benchmark prior. This corresponds to the case of no uncertainty, or the classical Bayesian environment. For $\epsilon$ sufficiently large, the uncertainty set of each agent $i$ coincides with the set of all probability distributions over $\Theta_{-i}$. This corresponds to the case of global uncertainty where the designer has no information whatsoever regarding agents' beliefs.

We write $U^\epsilon = \{U_i^\epsilon\}_{i\in\mathcal{I}}$ to denote the local uncertainty in the design environment. We implicitly assumed that the degree of local uncertainty is the same for all agents. This is for ease of exposition only; the model and the results can be readily extended to the case in which the degree of local uncertainty varies across agents. We do not need to assume that each agent's true belief about the other agents' types is independent of her own type. But as we shall see shortly, for the solution concept of robust incentive compatibility, which requires that the Bayesian incentive constraint be satisfied for every plausible belief of each type, it is without loss of generality to assume that each agent' true belief is independent of her own type.

Lemma 1 below presents a topological property of the uncertainty set $U_i^\epsilon$, which will be used in later analysis.

**Lemma 1.** *There is a finite set $E_i$ of extreme points of $U_i^\epsilon$, and $U_i^\epsilon = \text{Conv.Hull}\,(E_i)$.*

*Proof.* It is easy to see that $U_i^\epsilon$ is a nonempty, convex, and compact set in $\mathbb{R}^{|\Theta_{-i}|}$. Furthermore, by definition, $U_i^\epsilon$ is the bounded intersection of finitely many closed half-spaces in $\mathbb{R}^{|\Theta_{-i}|}$. Thus, $U_i^\epsilon$ is a convex polytope with a finite set of vertices, which is also the set of its extreme points. It follows from Minkowski's theorem (see Hiriart-Urruty and Lemaréchal (2004, Theorem 2.3.4)) that $U_i^\epsilon$ is the convex hull of its extreme points. □

## 2.3 Mechanisms and solution concepts

A mechanism consists of a set $M_i$ of messages for each agent $i$, a decision rule $q : \prod_{i\in\mathcal{I}} M_i \to \Delta\mathcal{K}$, and a transfer rule $t : \prod_{i\in\mathcal{I}} M_i \to \mathbb{R}^I$. The revelation principle holds for the solution concepts we consider in this paper (including robust incentive compatibility with respect to an uncertainty set), and we restrict attention to direct mechanisms where $M_i = \Theta_i$ for each agent $i$.[6] The agents are asked to simultaneously report their types to the designer. Based on the reported type profile, the decision rule $q$ specifies the outcome in $\Delta\mathcal{K}$ with $q^k$ representing the probability that alternative $k$ is chosen and the transfer rule $t_i$ specifies how much agent $i$ pays to the designer.

The mechanism design problem is to fix a solution concept and search for the mechanism that maximizes the designer's payoff (in some outcome consistent with the solution concept). A standard solution concept is that of a Bayesian Nash equilibrium—It requires that truth-telling constitute a Bayesian Nash equilibrium with respect to the benchmark prior.

**Definition 1.** *A direct mechanism $(q, t)$ is Bayesian incentive compatible with respect to the benchmark prior $\mu$ if for each agent $i$ and each type $\theta_i \in \Theta_i$,*

$$\sum_{\theta_{-i}\in\Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big)\, \mu_i(\theta_{-i}) \geq \sum_{\theta_{-i}\in\Theta_{-i}} \Big(q(\theta_i', \theta_{-i}) \cdot \theta_i - t_i(\theta_i', \theta_{-i})\Big)\, \mu_i(\theta_{-i})$$

*for any alternative type $\theta_i' \in \Theta_i$, and*

$$\sum_{\theta_{-i}\in\Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big)\, \mu_i(\theta_{-i}) \geq 0.$$

A mechanism $\Gamma$ that satisfies Bayesian incentive compatibility with respect to the benchmark prior $\mu$ is not robust to misspecification of agents' beliefs. When the agents' actual beliefs are not consistent with the benchmark prior $\mu$, truth-telling might not constitute a Bayesian Nash equilibrium in the mechanism $\Gamma$. To minimize the role of the assumption about agents' beliefs, a common approach is to adopt the stronger solution concept of dominant strategy incentive compatibility—It requires that truth-telling constitute a dominant strategy equilibrium.

**Definition 2.** *A direct mechanism $(q, t)$ is dominant strategy incentive compatible if for each agent $i$, each type $\theta_i \in \Theta_i$, and $\theta_{-i} \in \Theta_{-i}$,*

$$q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i}) \geq q(\theta_i', \theta_{-i}) \cdot \theta_i - t_i(\theta_i', \theta_{-i})$$

[6] Lopomo, Rigotti, and Shannon (2020) prove a version of the revelation principle in the case in which there exists uncertainty about agents' beliefs—If a social choice function can be robustly implemented by some mechanism, then it can also be robustly truthfully implemented in a direct mechanism.

*for any alternative type* $\theta_i' \in \Theta_i$*, and*

$$q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i}) \geq 0.$$

Evidently, the notion of dominant strategy incentive compatibility does not rely on any assumptions about agents' beliefs and is thus robust to any uncertainty about agents' beliefs. It applies to the situation where the designer does not have any reliable information about agents' beliefs. However, the notion of dominant strategy incentive compatibility is rather demanding; for many mechanism design problems, the class of dominant strategy mechanisms is quite small, and only includes mechanisms that are rather unattractive for the designer.

In many realistic settings, the designer may possess some partial information about agents' beliefs even if she cannot pin down the exact beliefs held by the agents. Particularly, in the case in which the designer faces only local uncertainty about agents' beliefs, she searches over mechanisms that are Bayesian incentive compatible with respect to the uncertainty set and thus are robust to local misspecification of agents' beliefs. The notion of robust incentive compatibility with respect to the uncertainty set guarantees that truth-telling constitutes a Bayesian Nash equilibrium that is robust to local uncertainty about agents' beliefs.

**Definition 3.** *A direct mechanism* $(q, t)$ *is robust incentive compatible with respect to the uncertainty set* $U^\epsilon$ *if, for each agent* $i$*, each type* $\theta_i \in \Theta_i$*, and each belief* $\nu_i \in U_i^\epsilon$*,*

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big) \nu_i(\theta_{-i}) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i', \theta_{-i}) \cdot \theta_i - t_i(\theta_i', \theta_{-i})\Big) \nu_i(\theta_{-i})$$

*for any alternative type* $\theta_i' \in \Theta_i$*, and*

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big) \nu_i(\theta_{-i}) \geq 0.$$

It follows from Lemma 1 that a direct mechanism is robust incentive compatible with respect to the uncertainty set if and only if the Bayesian incentive constraints hold for every extreme point of the uncertainty set. This observation leads to the following equivalent definition for the notion of robust incentive compatibility.

**Definition 4.** *A direct mechanism* $(q, t)$ *is robust incentive compatible with respect to* *the uncertainty set $U^\epsilon$ if, for each agent $i$, each type $\theta_i \in \Theta_i$, and each belief $\nu_i \in E_i$,*

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big)\, \nu_i(\theta_{-i}) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i', \theta_{-i}) \cdot \theta_i - t_i(\theta_i', \theta_{-i})\Big)\, \nu_i(\theta_{-i})$$

*for any alternative type $\theta_i' \in \Theta_i$, and*

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big)\, \nu_i(\theta_{-i}) \geq 0.$$

For simplicity of exposition, we add a dummy type $\theta_0$ for each agent $i \in \mathcal{I}$ that corresponds to not participating in the mechanism; let $q(\theta_0, \theta_{-i}) \cdot \theta_i = t_i(\theta_0, \theta_{-i}) = 0$ for all $\theta_i \in \Theta_i$ and $\theta_{-i} \in \Theta_{-i}$.

## 2.4 The designer's robust optimization problem

Given the local uncertainty about agents' beliefs, the designer chooses a robust incentive compatible mechanism that maximizes her expected revenue. The robust incentive compatible mechanism design problem can be formulated as the following robust optimization problem:

$$\max_{q(\cdot),\, t(\cdot)} \quad \sum_{\theta \in \Theta} \sum_{i \in \mathcal{I}} \mu(\theta)\, t_i(\theta) \tag{RIC}$$

subject to $\forall i \in \mathcal{I}$, $\forall \theta_i \in \Theta_i$, $\forall \nu_i \in E_i$, $\forall \theta_i' \in \{\Theta_i \setminus \{\theta_i\}\} \cup \{\theta_0\}$,

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big)\, \nu_i(\theta_{-i}) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i', \theta_{-i}) \cdot \theta_i - t_i(\theta_i', \theta_{-i})\Big)\, \nu_i(\theta_{-i}), \tag{1}$$

$$\forall \theta \in \Theta,\ \forall k \in \mathcal{K},\ q^k(\theta) \geq 0, \tag{2}$$

$$\forall \theta \in \Theta,\ \sum_{k \in \mathcal{K}} q^k(\theta) = 1. \tag{3}$$

The constraint (1) requires that the mechanism be robust incentive compatible with respect to the uncertainty set. The constraints (2) and (3) are standard feasibility constraints. Let $S^R(\epsilon)$ denote the value of the objective function of the robust optimization problem (RIC) at an optimum.

When $\epsilon = 0$, the uncertainty set of each agent $i$ is just a singleton set—the conditional probability distribution derived from the benchmark prior. The maximization problem (RIC) reduces to the following standard Bayesian mechanism design problem:

$$\max_{q(\cdot),\,t(\cdot)} \quad \sum_{\theta\in\Theta}\sum_{i\in\mathcal{I}} \mu(\theta)\, t_i(\theta) \tag{BIC}$$

$$\text{subject to} \quad \forall i\in\mathcal{I},\ \forall\theta_i\in\Theta_i,\ \forall\theta_i'\in\{\Theta_i\setminus\{\theta_i\}\}\cup\{\theta_0\},$$

$$\sum_{\theta_{-i}\in\Theta_{-i}} \Big(q(\theta_i,\theta_{-i})\cdot\theta_i - t_i(\theta_i,\theta_{-i})\Big)\,\mu_i(\theta_{-i})$$

$$\geq \sum_{\theta_{-i}\in\Theta_{-i}} \Big(q(\theta_i',\theta_{-i})\cdot\theta_i - t_i(\theta_i',\theta_{-i})\Big)\,\mu_i(\theta_{-i}),$$

$$\forall\theta\in\Theta,\ \forall k\in\mathcal{K},\ q^k(\theta)\geq 0,$$

$$\forall\theta\in\Theta,\ \sum_{k\in\mathcal{K}} q^k(\theta) = 1.$$

For sufficiently large $\epsilon$, the uncertainty set of each agent coincides with the set of all probability distributions over the other agents' types, and the maximization problem (RIC) reduces to the following standard dominant strategy mechanism design problem:

$$\max_{q(\cdot),\,t(\cdot)} \quad \sum_{\theta\in\Theta}\sum_{i\in\mathcal{I}} \mu(\theta)\, t_i(\theta) \tag{DIC}$$

$$\text{subject to} \quad \forall i\in\mathcal{I},\ \forall\theta_i\in\Theta_i,\ \forall\theta_i'\in\{\Theta_i\setminus\{\theta_i\}\}\cup\{\theta_0\},\ \forall\theta_{-i}\in\Theta_{-i},$$

$$q(\theta_i,\theta_{-i})\cdot\theta_i - t_i(\theta_i,\theta_{-i}) \geq q(\theta_i',\theta_{-i})\cdot\theta_i - t_i(\theta_i',\theta_{-i}),$$

$$\forall\theta\in\Theta,\ \forall k\in\mathcal{K},\ q^k(\theta)\geq 0,$$

$$\forall\theta\in\Theta,\ \sum_{k\in\mathcal{K}} q^k(\theta) = 1.$$

We write $S^B$ (resp. $S^D$) for the value of the objective function of the maximization problem (BIC) (resp. (DIC)) at an optimum.

Clearly, as the degree of local uncertainty increases, the uncertainty set becomes larger and the corresponding set of incentive constraints becomes more restrictive. Formally, if a mechanism $(q,t)$ is robust incentive compatible with respect to $U^\epsilon$, then it is also robust incentive compatible with respect to $U^{\epsilon'}$ for any $\epsilon' < \epsilon$. It follows that $S^R(\epsilon)$ (viewed as a function of $\epsilon$) is weakly decreasing in $\epsilon$. We record this observation as the following proposition.

**Proposition 1.** *$S^R(\epsilon)$ is weakly decreasing in $\epsilon$. In particular,*

$$S^D \leq S^R(\epsilon) \leq S^B$$

*for any $\epsilon$.*

# 3 Implementation

In this section, we characterize decision rules that are robustly implementable with respect to an uncertainty set.

**Definition 5.** *A decision rule $q$ is robustly implementable with respect to an uncertainty set if there exists a transfer scheme $t$ such that the mechanism $(q, t)$ is robust incentive compatible with respect to the given uncertainty set.*

## 3.1 Bayesian implementation

To characterize robust implementation with respect to an uncertainty set, it is instructive to first revisit the implementability of a decision rule in the standard Bayesian framework. Consider some $\nu = (\nu_1, \nu_2, \ldots, \nu_I)$, where $\nu_i \in E_i$ for each agent $i \in \mathcal{I}$.

**Definition 6.** *A decision rule $q$ is implementable with respect to $\nu$ if there exists a transfer scheme $t$ such that the mechanism $(q, t)$ is Bayesian incentive compatible with respect to $\nu$.*

The following lemma characterizes the implementability of a decision rule $q$ with respect to $\nu$.

**Lemma 1.** *(Rochet (1987)) A decision rule $q$ is implementable with respect to $\nu$ if and only if $q$ is interim cyclically monotone with respect to $\nu$, that is, for each agent $i \in \mathcal{I}$ and each sequence of types $(\theta_i^1, \theta_i^2, \ldots, \theta_i^n) \in \Theta_i^n$ with $\theta_i^n = \theta_i^1$, we have*

$$\sum_{t=1}^{n-1} \Big( \sum_{\theta_{-i} \in \Theta_{-i}} \left( q(\theta_i^t, \theta_{-i}) \cdot \theta_i^{t+1} \right) \nu_i(\theta_{-i}) - \sum_{\theta_{-i} \in \Theta_{-i}} \left( q(\theta_i^t, \theta_{-i}) \cdot \theta_i^t \right) \nu_i(\theta_{-i}) \Big) \leq 0.$$

In what follows, we pin down a particular class of transfer schemes that implement the decision rule $q$ with respect to $v$. Given a decision rule $q$ that is implementable with respect to $\nu$, consider the following optimization problem:

$$\max_{t_i(\cdot)} \quad \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \sum_{\theta_{-i} \in \Theta_{-i}} t_i(\theta_i, \theta_{-i})\, \nu_i(\theta_{-i}) \tag{Primal-$\nu_i$}$$

$$\text{subject to} \quad \forall \theta_i \in \Theta_i,\ \forall \theta_i' \in \{\Theta_i \setminus \{\theta_i\}\} \cup \{\theta_0\},$$

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i})\Big)\, \nu_i(\theta_{-i})$$

$$\geq \sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i', \theta_{-i}) \cdot \theta_i - t_i(\theta_i', \theta_{-i})\Big)\, \nu_i(\theta_{-i}).$$

We derive its dual minimization problem as follows:

$$\min_{\lambda^{\nu_i}(\cdot)} \quad \sum_{\theta_i \in \Theta_i} \sum_{\theta_i' \in \{\Theta_i \setminus \{\theta_i\}\} \cup \{\theta_0\}} \lambda^{\nu_i}(\theta_i, \theta_i')\, w^{\nu_i}(\theta_i, \theta_i') \tag{Dual-$\nu_i$}$$

$$\text{subject to} \quad \forall \theta_i \in \Theta_i,$$

$$\sum_{\theta_i' \in \{\Theta_i \setminus \{\theta_i\}\} \cup \{\theta_0\}} \lambda^{\nu_i}(\theta_i, \theta_i') - \sum_{\theta_i' \in \Theta_i \setminus \{\theta_i\}} \lambda^{\nu_i}(\theta_i', \theta_i) = \mu_i(\theta_i),$$

$$\forall \theta_i \in \Theta_i,\ \forall \theta_i' \in \{\Theta_i \setminus \{\theta_i\}\} \cup \{\theta_0\},$$

$$\lambda^{\nu_i}(\theta_i, \theta_i') \geq 0,$$

where

$$w^{\nu_i}(\theta_i, \theta_i') = \sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - q(\theta_i', \theta_{-i}) \cdot \theta_i\Big)\, \nu_i(\theta_{-i})$$

and $\lambda^{\nu_i}(\theta_i, \theta_i')$ is the multiplier on the incentive constraint in the optimization problem (Primal-$\nu_i$).

The dual problem (Dual-$\nu_i$) is a network flow problem on the following network $G(q, \nu_i)$ where

(1) the set of nodes is $\Theta_i \cup \{\theta_0\}$ (the node corresponding to the dummy type $\theta_0$ is the source),

(2) for any $\theta_i \in \Theta_i$ and $\theta_i' \in \{\Theta_i \setminus \{\theta_i\}\} \cup \{\theta_0\}$, $\theta_i' \to \theta_i$ is a directed edge with length

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(q(\theta_i, \theta_{-i}) \cdot \theta_i - q(\theta_i', \theta_{-i}) \cdot \theta_i\Big)\, \nu_i(\theta_{-i}),$$

(3) a path from the dummy type $\theta_0$ to type $\theta_i^t$ is a sequence $P = (\theta_0, \theta_i^1, \theta_i^2, \ldots, \theta_i^t)$

of nodes where $\theta_i^k \in \Theta_i$ for all $k$.

**Definition 7.** *Fix a decision rule $q$ that is implementable with respect to $\nu$. A shortest-path tree in the network $G(q, \nu_i)$ is the union of shortest paths from the source to all nodes.*

The optimization problem (Primal-$\nu_i$) is equivalent to determining the shortest-path tree in this network. At an optimum, edges on the shortest-path tree correspond to the binding constraints in (Primal-$\nu_i$).[7] Let $d(q, \nu_i, \theta_i)$ denote the length of the shortest path from the dummy type $\theta_0$ to type $\theta_i$ in the network $G(q, \nu_i)$. Standard arguments imply the following relationship between the length of the shortest path from the source to this node and the solution to the maximization problem (Primal-$\nu_i$):

$$d(q, \nu_i, \theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} t_i(\theta_i, \theta_{-i})\, \nu_i(\theta_{-i}). \tag{4}$$

Clearly, any payment scheme $t$ that satisfies (4) implements $q$ with respect to $\nu$. There are typically multiple payment schemes that fulfill this requirement, since (4) is a constraint on the interim expected payment rather than the ex post payment.

## 3.2 Robust implementation

Next, we return to the problem of robust implementation with respect to an uncertainty set. Clearly, if a decision rule is robustly implementable with respect to the uncertainty set, then the decision rule must be implementable with respect to any $\nu$ where $\nu_i \in E_i$ for each agent $i$. Say that $q$ is interim cyclically monotone with respect to an uncertainty $U^\epsilon$ if $q$ is interim cyclically monotone with respect to any $v$ where $v_i \in E_i$ for each agent $i$. The analysis in the previous subsection immediately implies a necessary condition for robust implementability, that is, if a decision rule $q$ is robustly implementable with respect to an uncertainty set, then $q$ is interim cyclically monotone with respect to the uncertainty set.

In what follows, we introduce some stronger conditions to ensure that a decision rule is robustly implementable with respect to some uncertainty set, building on the

[7] Readers unfamiliar with network flows may consult Ahuja, Magnanti, and Orlin (1993) and Vohra (2011). Vohra (2011, pp. 110-112) explicitly provides the network interpretation of the expected revenue maximization problems that we adopt in the current paper, and Vohra (2011, Chapter 6) applies this technique to study a variety of revenue maximization problems.

notion of the shortest-path tree and the particular class of transfer schemes in Section 3.1.

Suppose that $q$ is implementable with respect to any $\nu$ where $\nu_i \in E_i$ for each agent $i$. For each $\nu$, we focus on payment schemes that solve the corresponding maximization problem (Primal-$\nu_i$). This leads to a natural sufficient condition for a decision rule to be robustly implementable with respect to the uncertainty set. If there exists a payment scheme $t$ such that (4) holds for any $\nu$ where $\nu_i \in E_i$ for each agent $i$, then the payment scheme $t$ robustly implements $q$ with respect to the uncertainty set. The following formalizes this observation using a rank condition. Let $V(\theta_i) \in \mathbb{R}^{L\times|\Theta_{-i}|}$ denote the matrix where each row vector $v_i^l$ corresponds to an extreme point of the uncertainty set $U_i^\epsilon$. Let $\beta(\theta_i) \in \mathbb{R}^L$ denote the column vector that represents the length of the shortest path from the dummy type to type $\theta_i$ for each network $G(q, \nu_i)$. That is,

$$V(\theta_i) := \begin{pmatrix} \nu_i^1 \\ \vdots \\ \nu_i^L \end{pmatrix} \text{ and } \beta(\theta_i) := \begin{pmatrix} d(q, \nu_i^1, \theta_i) \\ \vdots \\ d(q, \nu_i^L, \theta_i) \end{pmatrix}.$$

Evidently, there exists a transfer scheme $t$ such that $(q, t)$ is robust incentive compatible with respect to the uncertainty set when the following condition holds:

$$\text{rank } V(\theta_i) = \text{rank } (V(\theta_i), \beta(\theta_i)), \ \forall i \in \mathcal{I}, \ \forall \theta_i \in \Theta_i. \tag{5}$$

**Proposition 2.** *If*

*(1) $q$ is interim cyclically monotone with respect to $U^\epsilon$, and*

*(2) for each agent $i$ and each type $\theta_i$, rank $(V(\theta_i), \beta(\theta_i)) =$ rank $V(\theta_i)$,*

*then $q$ is robustly implementable with respect to the uncertainty set $U^\epsilon$.*

**Remark 1.** The rank condition in Proposition 2 is vacuous in the case of Bayesian incentive compatibility and dominant strategy incentive compatibility, tying our results to the classical results that characterize implementable decision rules under these alternative solution concepts. First consider Bayesian incentive compatibility with respect to $\mu$. In this case, $U_i^\epsilon = \{\mu_i\}$ for each agent $i$, $L = 1$, and rank $V(\theta_i) =$ rank $(V(\theta_i), \beta(\theta_i)) = 1$ for each agent $i$ and each type $\theta_i \in \Theta_i$. Next consider dominant strategy incentive compatibility. In this case, $U_i^\epsilon = \Delta\Theta_{-i}$ for each agent $i$, $V(\theta_i)$ is the identity matrix and rank $V(\theta_i) = \text{rank}\ (V(\theta_i), \beta(\theta_i)) = |\Theta_{-i}|$ for each agent $i$ and each type $\theta_i \in \Theta_i$.

Corollary 1 below presents a class of environments in which the rank condition is satisfied.

**Definition 8.** *A decision rule $q$ satisfies the uniform shortest-path tree property with respect to $U^\epsilon$ if for each agent $i \in \mathcal{I}$, there is the same shortest-path tree for all networks $G(q, \nu_i)$ where $\nu_i \in E_i$.*

**Corollary 1.** *If $q$ satisfies the uniform shortest-path tree property with respect to $U^\epsilon$, then $q$ is robustly implementable with respect to the uncertainty set $U^\epsilon$.*

*Proof.* It suffices to show that the rank condition in Proposition 2 is satisfied. For each agent $i$ and each type $\theta_i \in \Theta_i$, we write $(\theta_0, \theta_i^1, \theta_i^2, \ldots, \theta_i^t, \theta_i)$ for the (uniform) shortest path in all networks $G(q, \nu_i)$ where $\nu_i \in E_i$. Then for any $\nu_i \in E_i$,

$$\begin{aligned} d(q, \nu_i, \theta_i) = & \sum_{\theta_{-i} \in \Theta_{-i}} \left( q(\theta_i^1, \theta_{-i}) \cdot \theta_i^1 \right) \nu_i(\theta_{-i}) \\ & + \sum_{\theta_{-i} \in \Theta_{-i}} \left( (q(\theta_i^2, \theta_{-i}) - q(\theta_i^1, \theta_{-i})) \cdot \theta_i^2 \right) \nu_i(\theta_{-i}) \\ & + \ldots \\ & + \sum_{\theta_{-i} \in \Theta_{-i}} \left( (q(\theta_i^t, \theta_{-i}) - q(\theta_i^{t-1}, \theta_{-i})) \cdot \theta_i^t \right) \nu_i(\theta_{-i}) \\ & + \sum_{\theta_{-i} \in \Theta_{-i}} \left( (q(\theta_i, \theta_{-i}) - q(\theta_i^t, \theta_{-i})) \cdot \theta_i \right) \nu_i(\theta_{-i}). \end{aligned}$$

Clearly, the rank condition in Proposition 2 is satisfied. □

As an illustration, consider the classical environment of single-unit auction. Each agent $i \in \mathcal{I}$ has $M$ possible valuations for the good. For notational simplicity, we assume that the set of possible valuations is the same for each agent: $\Theta_i = \{\theta^1, \theta^2, \ldots, \theta^M\}$, where $\theta^m - \theta^{m-1} = \gamma$ for each $m = 2, 3, \ldots, M$ for some $\gamma > 0$. A direct mechanism is denoted by $(q, t)$, where $q_i \in [0, 1]$ denotes the probability that agent $i$ obtains the good, and $t_i$ denotes the payment to the seller.

We write

$$Q^{\nu_i}(\theta^m) = \sum_{\theta_{-i} \in \Theta_{-i}} q_i(\theta^m, \theta_{-i})\, \nu_i(\theta_{-i})$$

for the interim probability that agent $i$ obtains the good (with respect to $\nu_i$). We say that $q$ is interim increasing with respect to $U^\epsilon$ if for each agent $i \in \mathcal{I}$ and each belief $\nu_i \in E_i$, $Q^{\nu_i}(\theta^m)$ is increasing in $m$. For any $q$ that is interim increasing with respect to $U^\epsilon$, $q$ is interim cyclically monotone with respect to $U^\epsilon$, and the shortest-path tree in each network $G(q, \nu_i)$ is well-defined. The condition in Corollary 1 is satisfied in this environment—For any agent $i \in \mathcal{I}$, the types are completely ordered via a single path. We omit the proof of the following lemma.

**Lemma 2.** *Fix any $q$ that is interim increasing with respect to $U^\epsilon$. The shortest path from the dummy type $\theta_0$ to any type $\theta^m$ in each network $G(q, \nu_i)$ is $(\theta_0, \theta^1, \theta^2, \ldots, \theta^m)$.*

We have the following necessary and sufficient condition for a decision rule $q$ to be robustly implementable with respect to the uncertainty set.

**Corollary 2.** *In the single-unit auction environment, a decision rule $q$ is robustly implementable with respect to the uncertainty set $U^\epsilon$ if and only if $q$ is interim increasing with respect to the uncertainty set $U^\epsilon$.*

# 4 Robustness to local uncertainty for fixed $\epsilon$

In this section, we focus on the designer's robust optimization problem (RIC) for any fixed $\epsilon$. We identify necessary and sufficient conditions under which requiring robustness to local uncertainty about agents' beliefs is without loss of revenue for the designer.

## 4.1 Robustness to local uncertainty without loss of revenue

Our analysis is based on the uniform shortest-path tree property with respect to an uncertainty set. Fix a decision rule $q$ that satisfies the uniform shortest-path tree property with respect to an uncertainty set $U^\epsilon$. Since for each agent $i$, the shortest-path tree is uniform across all networks $G(q, \nu_i)$ where $v_i \in E_i$, for ease of notation, we drop the dependence of the shortest-path tree on $\nu_i$. We represent the uniform shortest-path tree of agent $i$ by a partial order $\succ_i$ on $i$'s types and its transitive closure $\succ_i^+$; we write $\theta_i \succ_i \theta_i^t \succ_i \cdots \succ_i \theta_i^2 \succ_i \theta_i^1 \succ_i \theta_0$ if the uniform shortest path from the dummy type $\theta_0$ to type $\theta_i$ is $(\theta_0, \theta_i^1, \theta_i^2, \ldots, \theta_i^t, \theta_i)$. We write $\theta_i' \succeq_i^+ \theta_i$ if $\theta_i' \succ_i^+ \theta_i$ or $\theta_i' = \theta_i$. If $\theta_i \succ_i \theta_i'$, we sometimes denote $\theta_i'$ by $\theta_i^-$.

We first present a sufficient condition under which the local uncertainty about agents' beliefs will not decrease the expected revenue for the designer.

**Proposition 3.** *If there exists some optimal decision rule $q^*$ from (BIC) such that $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$, then requiring robustness to local uncertainty about agents' beliefs is without loss of revenue for the designer, that is,*

$$S^R(\epsilon) = S^B.$$

*Proof.* We fix the decision rule $q^*$ and work with the payment scheme. Step 1 considers the robust optimization problem. We work with the maximization problem (RIC-P) and derive its dual (RIC-D). Denote by $V_{RIC-P}$ (resp. $V_{RIC-D}$) the value of the objective function of the program (RIC-P) (resp. (RIC-D)) at an optimum. Step 2 considers the optimal Bayesian mechanism design problem (BIC-P), and derives its dual (BIC-D). Denote by $V_{BIC-P}$ (resp. $V_{BIC-D}$) the value of the objective function of the program (BIC-P) (resp. (BIC-D)) at an optimum. Step 3 proceeds to show that $V_{RIC-D} \geq V_{BIC-D}$. It follows from the duality theorem in linear programming that $V_{RIC-P} = V_{RIC-D} \geq V_{BIC-D} \geq V_{BIC-P} \geq V_{RIC-P}$, which then implies $V_{RIC-P} = V_{BIC-P}$. Thus, requiring robustness to local uncertainty about agents' beliefs is without loss of revenue.[8]

**Step 1.** Since $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$, by Corollary 1, $q^*$ is robustly implementable with respect to the uncertainty set $U^\epsilon$. Given $q^*$, we solve for the optimal transfer scheme $t$ that maximizes the designer's expected revenue subject to the robust incentive constraints. It is without loss of generality to solve this problem by solving $t_i$ for each agent $i$ separately. That is,

$$\begin{aligned}
\max_{t(\cdot)} \quad & \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \sum_{\theta_{-i} \in \Theta_{-i}} t_i(\theta_i, \theta_{-i})\, \mu_i(\theta_{-i}) \\
\text{subject to} \quad & \forall \theta_i \in \Theta_i,\ \forall \nu_i \in E_i,\ \forall \theta_i' \in \Theta_i \cup \{\theta_0\}, \\
& \sum_{\theta_{-i} \in \Theta_{-i}} \Big( q^*(\theta_i, \theta_{-i}) \cdot \theta_i - t_i(\theta_i, \theta_{-i}) \Big)\, \nu_i(\theta_{-i}) \\
& \geq \sum_{\theta_{-i} \in \Theta_{-i}} \Big( q^*(\theta_i', \theta_{-i}) \cdot \theta_i - t_i(\theta_i', \theta_{-i}) \Big)\, \nu_i(\theta_{-i}).
\end{aligned}$$

[8]One can also show this result directly, without using the duality approach.

By Lemma 1, $\mu_i$ can be expressed as a convex combination of the extreme points of $U^\epsilon$. This further implies that the objective function in the above maximization problem for each agent $i$ can be expressed as a convex combination of the objective functions in the subproblems (Primal-$\nu_i$) for all $\nu_i \in E_i$. Since for each subproblem (Primal-$\nu_i$) (corresponding to each $\nu_i \in E_i$), it suffices to only consider constraints that correspond to edges on the uniform shortest-path tree, we conclude that in the above maximization problem for each agent $i$, it suffices to only consider constraints that correspond to edges on the uniform shortest-path tree. Thus, the maximization problem for each agent $i$ reduces to the following:

$$\max_{t_i(\cdot)} \quad \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \sum_{\theta_{-i} \in \Theta_{-i}} t_i(\theta_i, \theta_{-i})\, \mu_i(\theta_{-i}) \tag{RIC-P}$$

$$\text{subject to} \quad \forall \theta_i \in \Theta_i,\ \forall \nu_i \in E_i,$$

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(t_i(\theta_i, \theta_{-i}) - t_i(\theta_i^-, \theta_{-i})\Big)\, \nu_i(\theta_{-i}) \le w^{\nu_i}(\theta_i, \theta_i^-),$$

where

$$w^{\nu_i}(\theta_i, \theta_i^-) = \sum_{\theta_{-i} \in \Theta_{-i}} \Big(q^*(\theta_i, \theta_{-i}) \cdot \theta_i - q^*(\theta_i^-, \theta_{-i}) \cdot \theta_i\Big)\, \nu_i(\theta_{-i}).$$

It follows from standard arguments that $V_{RIC-P}$ is finite.

We derive its dual minimization problem (RIC-D) as follows:

$$\min_{\lambda^{RIC}(\cdot)} \quad \sum_{\theta_i \in \Theta_i} \sum_{\nu_i \in E_i} \lambda^{RIC}(\theta_i, \nu_i)\, w^{\nu_i}(\theta_i, \theta_i^-) \tag{RIC-D}$$

$$\text{subject to} \quad \forall \theta_i \in \Theta_i,\ \forall \theta_{-i} \in \Theta_{-i},$$

$$\sum_{\nu_i \in E_i} \lambda^{RIC}(\theta_i, \nu_i)\, \nu_i(\theta_{-i}) - \sum_{\theta_i' : \theta_i' \succ_i \theta_i} \sum_{\nu_i \in E_i} \lambda^{RIC}(\theta_i', \nu_i)\, \nu_i(\theta_{-i}) = \mu(\theta), \tag{6}$$

$$\forall \theta_i \in \Theta_i,\ \forall \nu_i \in E_i,\ \lambda^{RIC}(\theta_i, \nu_i) \ge 0,$$

where $\lambda^{RIC}(\theta_i, \nu_i)$ is the multiplier on the corresponding incentive constraint in (RIC-P).

**Step 2.** Next, we consider the Bayesian mechanism design problem and solve $t_i$ for each agent $i$ separately. Since $q^*$ satisfies the uniform shortest-path tree property with respect to $U^\epsilon$, the Bayesian revenue maximization problem with respect to the benchmark prior reduces to the following:

$$\max_{t_i(\cdot)} \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \sum_{\theta_{-i} \in \Theta_{-i}} t_i(\theta_i, \theta_{-i})\, \mu_i(\theta_{-i}) \tag{BIC-P}$$

subject to $\forall \theta_i \in \Theta_i$,

$$\sum_{\theta_{-i} \in \Theta_{-i}} \Big(t_i(\theta_i, \theta_{-i}) - t_i(\theta_i^-, \theta_{-i})\Big)\, \mu_i(\theta_{-i}) \leq w^{\mu_i}(\theta_i, \theta_i^-).$$

where

$$w^{\mu_i}(\theta_i, \theta_i^-) = \sum_{\theta_{-i} \in \Theta_{-i}} \Big(q^*(\theta_i, \theta_{-i}) \cdot \theta_i - q^*(\theta_i^-, \theta_{-i}) \cdot \theta_i\Big)\, \mu_i(\theta_{-i}).$$

We derive its dual minimization problem (BIC-D) as follows:

$$\min_{\lambda^{BIC}(\cdot)} \sum_{\theta_i \in \Theta_i} \lambda^{BIC}(\theta_i)\, w^{\mu_i}(\theta_i, \theta_i^-) \tag{BIC-D}$$

subject to $\forall \theta_i \in \Theta_i$,

$$\lambda^{BIC}(\theta_i) - \sum_{\theta_i' :\, \theta_i' \succ_i \theta_i} \lambda^{BIC}(\theta_i') = \mu_i(\theta_i),$$

$$\forall \theta_i \in \Theta_i,\ \lambda^{BIC}(\theta_i) \geq 0,$$

where $\lambda^{BIC}(\theta_i)$ is the multiplier on the corresponding incentive constraint in the maximization problem (BIC-P).

**Step 3.** We proceed to show that $V_{RIC-D} \geq V_{BIC-D}$. By inspecting the constraints and the objective functions in the two minimization problems (RIC-D) and (BIC-D), we know that for any feasible dual variables $\lambda^{RIC}(\cdot)$ with corresponding value of the objective function in the minimization problem (RIC-D), the following feasible dual variables $\lambda^{BIC}(\cdot)$ achieves the same value for the objective function of the minimization problem (BIC-D):[9]

$$\lambda^{BIC}(\theta_i) = \frac{\sum_{\nu_i \in E_i} \lambda^{RIC}(\theta_i, \nu_i)\, \nu_i(\theta_{-i})}{\mu_i(\theta_{-i})}.$$

[9]The RHS is constant in $v_{-i}$ for any feasible dual variables $\lambda^{RIC}(\cdot)$. By induction, we can derive the following from (6):

$$\frac{\sum_{\nu_i \in E_i} \lambda^{RIC}(\theta_i, \nu_i)\, \nu_i(\theta_{-i})}{\mu_i(\theta_{-i})} = \sum_{\theta_i' :\, \theta_i' \succeq_i^+ \theta_i} \mu_i(\theta_i'),\ \forall \theta_i \in \Theta_i,\ \forall \theta_{-i} \in \Theta_{-i}.$$

Thus, we can conclude that $V_{RIC-D} \geq V_{BIC-D}$.

It follows from the duality theorem in linear programming that $V_{RIC-P} = V_{RIC-D} \geq V_{BIC-D} \geq V_{BIC-P} \geq V_{RIC-P}$, which further implies $V_{RIC-P} = V_{BIC-P}$. Since $q^*$ is the optimal decision rule of the Bayesian mechanism design problem (BIC), we conclude that $S^R(\epsilon) = S^B$. □

Next, we show that the sufficient condition in Proposition 3 is also necessary to guarantee no revenue loss for requiring robustness to local uncertainty about agents' beliefs. For any decision rule $q$ that is implementable with respect to the benchmark prior, let $V_{BIC-q}$ be the highest expected revenue by maximizing over transfer schemes subject to the corresponding (BIC) constraints. For any decision rule $q$ that is robustly implementable with respect to the uncertainty set, let $V_{RIC-q}$ be the highest expected revenue by maximizing over transfer schemes subject to the corresponding (RIC) constraints.

**Proposition 4.** *If $S^R(\epsilon) = S^B$, then there exists some optimal decision rule $q^*$ from (BIC) such that $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$.*

*Proof.* We first show that if $S^R(\epsilon) = S^B$, then there exists some optimal decision rule $q^*$ from (BIC) such that $q^*$ is robustly implementable with respect to the uncertainty set. Suppose to the contrary, there does not exist such a decision rule. In other words, if $q$ is robustly implementable, then $q$ is not optimal for (BIC). In particular, the optimal decision rule $\tilde{q}$ from (RIC) is not optimal for (BIC). It follows that

$$S^R(\epsilon) = V_{RIC-\tilde{q}} \leq V_{BIC-\tilde{q}} < S^B,$$

which contradicts that $S^R(\epsilon) = S^B$.

Let $\mathcal{Q}$ be the collection of all decision rules that are optimal for (BIC) and are also robustly implementable with respect to the uncertainty set. It follows from the above that $\mathcal{Q}$ is nonempty. In what follows, we show that there exists $q^* \in \mathcal{Q}$ that satisfies the uniform shortest-tree property with respect to $U^\epsilon$.

Suppose to the contrary, for any $q \in \mathcal{Q}$, there exists some agent $j \in \mathcal{I}$ and some belief $\tilde{\nu}_j \in E_j$ such that the shortest-path tree in the network $G(q, \tilde{\nu}_j)$ differs from that in the network $G(q, \mu_j)$. Then there exists some type $\tilde{\theta}_j \in \Theta_j$ such that the shortest path from the dummy type $\theta_0$ to $\tilde{\theta}_j$ in the network $G(q, \tilde{\nu}_j)$ differs from that in the network $G(q, \mu_j)$.

Given the decision rule $q$, let $t^B(\cdot)$ be the optimal transfer with respect to the corresponding (BIC) constraints. Because $q$ is an optimal decision rule for (BIC), $S^B$ can be achieved by maximizing over the payment schemes given $q$. We have

$$\begin{aligned} S^B &= \sum_{i \in \mathcal{I}} \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \Big( \sum_{\theta_{-i} \in \Theta_{-i}} t_i^B(\theta_i, \theta_{-i}) \, \mu_i(\theta_{-i}) \Big) \\ &= \sum_{i \in \mathcal{I}} \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \, d(q, \mu_i, \theta_i). \end{aligned}$$

On the other hand, given this decision rule $q$, let $t^R(\cdot)$ be the optimal transfer with respect to the corresponding (RIC) constraints. Therefore,

$$\begin{aligned} V_{RIC-q} &= \sum_{i \in \mathcal{I}} \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \Big( \sum_{\theta_{-i} \in \Theta_{-i}} t_i^R(\theta_i, \theta_{-i}) \, \mu_i(\theta_{-i}) \Big) \\ &= \sum_{i \in \mathcal{I}} \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \sum_{\nu_i \in E_i} \alpha(\nu_i) \Big( \sum_{\theta_{-i} \in \Theta_{-i}} t_i^R(\theta_i, \theta_{-i}) \, \nu_i(\theta_{-i}) \Big) \\ &\leq \sum_{i \in \mathcal{I}} \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \sum_{\nu_i \in E_i} \alpha(\nu_i) \, d(q, \nu_i, \theta_i), \end{aligned}$$

where the second line follows from Lemma 1—there exist weights $\alpha(\cdot) \in [0, 1]$ such that $\sum_{\nu_i \in E_i} \alpha(\nu_i) = 1$ and $\mu_i(\theta_{-i}) = \sum_{\nu_i \in E_i} \alpha(\nu_i) \, \nu_i(\theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$. Furthermore, since $\mu$ is in the interior of the uncertainty set, we can choose the weights such that $\alpha(\tilde{\nu}_j) > 0$. We have the last line as an inequality since $t_i^R(\cdot)$ may not be optimal for the subproblems (Primal-$\nu_i$) with respect to each $\nu_i$.

For each agent $i \in \mathcal{I}$, we let $SP(q, \nu_i, \theta_i)$ denote the shortest path from the dummy type $\theta_0$ to type $\theta_i$ in the network $G(q, \nu_i)$. For each agent $i \in \mathcal{I}$, each belief $\nu_i \in E_i$, and each type $\theta_i \in \Theta_i$, by the definition of the shortest path in the network $G(q, \nu_i)$, we have

$$d(q, \nu_i, \theta_i) \leq \sum_{(\theta'_i, \theta''_i) \in SP(q, \mu_i, \theta_i)} w^{\nu_i}(\theta''_i, \theta'_i)$$

Furthermore, since $SP(q, \tilde{\nu}_j, \tilde{\theta}_j) \neq SP(q, \mu_j, \tilde{\theta}_j)$, we have

$$d(q, \tilde{\nu}_j, \tilde{\theta}_j) < \sum_{(\theta'_j, \theta''_j) \in SP(q, \mu_j, \tilde{\theta}_j)} w^{\tilde{\nu}_j}(\theta''_j, \theta'_j).$$

Therefore, we have

$$
\begin{aligned}
V_{RIC-q} &\leq \sum_{i\in\mathcal{I}} \sum_{\theta_i\in\Theta_i} \mu_i(\theta_i) \sum_{\nu_i\in E_i} \alpha(\nu_i)\, d(q,\nu_i,\theta_i) \\
&< \sum_{i\in\mathcal{I}} \sum_{\theta_i\in\Theta_i} \mu_i(\theta_i) \sum_{\nu_i\in E_i} \alpha(\nu_i) \sum_{(\theta_i',\theta_i'')\in SP(q,\mu_i,\theta_i)} w^{\nu_i}(\theta_i'',\theta_i') \\
&= \sum_{i\in\mathcal{I}} \sum_{\theta_i\in\Theta_i} \mu_i(\theta_i) \sum_{(\theta_i',\theta_i'')\in SP(q,\mu_i,\theta_i)} w^{\mu_i}(\theta_i'',\theta_i') \\
&= \sum_{i\in\mathcal{I}} \sum_{\theta_i\in\Theta_i} \mu_i(\theta_i)\, d(q,\mu_i,\theta_i) \\
&= S^B,
\end{aligned}
$$

where the third line follows from the definition of $w^{\nu_i}(\cdot)$ and $\mu_i(\theta_{-i}) = \sum_{\nu_i\in E_i} \alpha(\nu_i)\,\nu_i(\theta_{-i})$ for all $\theta_{-i}\in\Theta_{-i}$.

From the above, we conclude that for any $q\in\mathcal{Q}$, we have $V_{RIC-q} < S^B$. Clearly, for a decision rule $\tilde{q}$ that is robustly implementable with respect to the uncertainty set $U^\epsilon$ but is not optimal for (BIC), we also have

$$V_{RIC-\tilde{q}} \leq V_{BIC-\tilde{q}} < S^B.$$

Thus, it must be that $S^R(\epsilon) < S^B$. We arrive at a contradiction. □

**Theorem 1.** *$S^R(\epsilon) = S^B$ if and only if there exists some optimal decision rule $q^*$ from (BIC) such that $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$.*

Theorem 1 follows from Proposition 3 and Proposition 4.

## 4.2 Applications

**Environments in which $S^B = S^R(\epsilon) = S^D$ for any $\epsilon$.** The uniform shortest-path tree property with respect to an uncertainty set is of interest because a number of resource allocation problems satisfy this condition. Particularly, the uniform shortest-path tree property with respect to an uncertainty set (for any $\epsilon$) is satisfied in environments with one-dimensional types. This fits many classical applications of mechanism design, including single-unit auction (e.g., Myerson (1981)), public good (e.g., Mailath and Postlewaite (1990)), and standard bilateral trade (e.g., Myerson and Satterthwaite (1983)). This property also holds in multi-unit auctions with homogeneous or heterogeneous goods, combinatorial auctions and the like, as long as the agents' private values are one-dimensional. In this case, the types are linearly ordered via a single path. The uniform shortest-path tree property with respect to an uncertainty set (for any $\epsilon$) can also be satisfied in some multi-dimensional environments such as the multi-unit auction with capacity-constrained bidders (see Malakhov and Vohra (2009)). In this case, the types are located on different paths and are only partially ordered. In all these environments, by Theorem 1, we have $S^B = S^R(\epsilon) = S^D$ for any $\epsilon$.

**Environments in which $S^B = S^R(\epsilon) > S^D$ for a range of $\epsilon$ values.** In some environments, the uniform shortest-path tree property with respect to an uncertainty set can be satisfied for smaller degrees of local uncertainty even if it does not hold for larger degrees of local uncertainty. Thus, by Theorem 1, while $S^D < S^B$, we have $S^R(\epsilon) = S^B$ for a range of $\epsilon$ values. This highlights the appealing features of the robust optimization approach—While the designer would get a strictly less expected revenue from the optimal dominant strategy mechanism, she could employ the robust optimization approach which guards against misspecification of agents' beliefs without any loss of revenue for the designer. In what follows, we revisit Example 1 and illustrate this using the bilateral trade model with ex ante unidentified traders.

Consider the problem of designing a trading platform for two traders, $A$ and $B$, with the goal of maximizing intermediation profit. Each trader can buy or (short) sell one unit of the asset and has private information about her valuation for the good. The platform cannot hold inventory (ex post market-clearing is imposed). The platform has the following estimate of the distribution $\mu$ of the traders' types:

| | $\theta_B^1 = \frac{1}{3}$ | $\theta_B^2 = 1$ |
|---|---|---|
| $\theta_A^1 = 0$ | $\frac{1}{12}$ | $\frac{1}{24}$ |
| $\theta_A^2 = \frac{2}{3}$ | $\frac{7}{12}$ | $\frac{7}{24}$ |

Based on the true distribution of the agents' types and the uncertainty set of the agents' possible beliefs, the platform chooses trading mechanisms $(q, t_A, t_B)$ to maximize the expected profit. For each reported type profile $(\theta_A, \theta_B)$, $q(\theta_A, \theta_B) \in [-1, 1]$ is the number of units agent A buys from agent B, $t_i(\theta_A, \theta_B) \in \mathbb{R}$ is the payment from agent $i$ to the platform. Agent A's utility from purchasing $q$ units of the good and paying a transfer $t_A$ is $q\theta_A - t_A$, and agent B's utility from selling $q$ units of the good and paying a transfer $t_B$ is $-q\theta_B - t_B$.[10]

**Proposition 5.** *For the bilateral trade environment with ex ante unidentified traders, (1) $S^B > S^D$, and (2) $S^R(\epsilon) = S^B$ if and only if $\epsilon \leq \frac{1}{6}$.*

The following is the unique optimal decision rule of the Bayesian mechanism design problem (BIC)—Trader $A$ buys 1 unit from trader $B$ if the type profile is $(\frac{2}{3}, \frac{1}{3})$, and sells 1 unit to trader $B$ otherwise. Since $q^*$ is also implementable in dominant strategies, $q^*$ is robustly implementable with respect to the uncertainty set $U^\epsilon$ for any $\epsilon$.

Let $G(q^*, \theta_B^1)$ (resp. $G(q^*, \theta_B^2)$) denote the network corresponding to the decision rule $q^*$ and the degenerate belief that trader $B$'s value is $\theta_B^1$ (resp. $\theta_B^2$). For trader A, the shortest path from the dummy type $\theta_0$ to type $\theta_A^1$ is $(\theta_0, \theta_A^1)$ in the network $G(q^*, \theta_B^1)$, and is $(\theta_0, \theta_A^2, \theta_A^1)$ in the network $G(q^*, \theta_B^2)$.[11] It follows from Theorem 1 that $S^B > S^D$. It is straightforward to show that $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$ if and only if $0 \leq \epsilon \leq \frac{1}{6}$. It follows from Theorem 1 that $S^R(\epsilon) = S^B$ if and only if $0 \leq \epsilon \leq \frac{1}{6}$. The detailed calculation is contained in the Appendix A.

## 5 Robustness to local uncertainty for $\epsilon \to 0$

Despite the reliance on strong common knowledge assumptions, the Bayesian approach is widely adopted in the mechanism design literature. A common (albeit informal) defense of the Bayesian approach often goes as follows: While the designer understands that the Bayesian model is an approximation at best, the validity of the Bayesian approach could be justified by numerous historical transactions. In this section, we employ the robust optimization approach to formalize this rationale. In the robust optimization framework, the degree of local uncertainty $\epsilon$ diminishes as the number of historical transactions increases, and a multitude of historical transactions corresponds to the scenario where $\epsilon$ tends towards zero. In what follows, we compare the Bayesian

[10]Perhaps surprisingly, this is not a one-dimensional environment; see Börgers (2015, Section 5.6). This is because (1) if a trader is selling, then a higher value is the "higher type" and (2) if a trader is buying, then a lower value is the "higher type".

[11]In the network $G(q^*, \theta_B^1)$, the length of the path $(\theta_0, \theta_A^1)$ is 0 and the length of the path $(\theta_0, \theta_A^2, \theta_A^1)$ is $\frac{2}{3}$. In the network $G(q^*, \theta_B^2)$, the length of the path $(\theta_0, \theta_A^1)$ is 0 and the length of the path $(\theta_0, \theta_A^2, \theta_A^1)$ is $-\frac{2}{3}$.

approach and the robust optimization approach in the limit case in which $\epsilon$ converges to zero.

## 5.1 $S^B = S^R(\epsilon)$ for sufficiently small $\epsilon$

To rationalize the Bayesian approach, we might hope that, for any environment, as long as the degree of local uncertainty is sufficiently small, requiring robustness to local uncertainty of agents' beliefs is without loss of revenue. Formally, we investigate whether there exists a threshold $\bar{\epsilon} > 0$ such that $S^B = S^R(\epsilon)$ as long as $\epsilon \leq \bar{\epsilon}$. By Theorem 1, it suffices to check whether there exists an optimal decision rule of the Bayesian mechanism design problem (BIC) that satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^{\bar{\epsilon}}$.

The following example demonstrates that such a threshold does not always exist.

**Example 2.** *We consider the setting of bilateral trade with ex ante unidentified traders (the same setting as in Section 4.2). The platform has the following estimate of the distribution $\mu$ of the traders' types:*

| | $\theta_B^1 = \frac{1}{3}$ | $\theta_B^2 = 1$ |
|---|---|---|
| $\theta_A^1 = 0$ | $\frac{1}{4}$ | $\frac{1}{4}$ |
| $\theta_A^2 = \frac{2}{3}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

*We first consider the Bayesian mechanism design problem (BIC). It is straightforward to calculate that the following is the optimal decision rule:*

$$q^*(\theta_A^1, \theta_B^1) = -1, \qquad q^*(\theta_A^1, \theta_B^2) = -1,$$
$$q^*(\theta_A^2, \theta_B^1) = 1, \qquad q^*(\theta_A^2, \theta_B^2) = -1.$$

*We focus on trader $A$, and consider the shortest-path tree in the network $G(q^*, \nu_A)$ for some belief $\nu_A$. The length of the path $(\theta_0, \theta_A^1)$ is*

$$\left(\theta_A^1\, q^*(\theta_A^1, \theta_B^1)\right) \nu_A(\theta_B^1) + \left(\theta_A^1\, q^*(\theta_A^1, \theta_B^2)\right) \nu_A(\theta_B^2) = 0,$$

*and the length of the path $(\theta_0, \theta_A^2, \theta_A^1)$ is*

$$\left(\theta_A^2\, q^*(\theta_A^2, \theta_B^1)\right) \nu_A(\theta_B^1) + \left(\theta_A^2\, q^*(\theta_A^2, \theta_B^2)\right) \nu_A(\theta_B^2)$$

$$+\Big(\theta_A^1\, q^*(\theta_A^1,\theta_B^1) - \theta_A^1\, q^*(\theta_A^2,\theta_B^1)\Big)\, \nu_A(\theta_B^1) + \Big(\theta_A^1\, q^*(\theta_A^1,\theta_B^2) - \theta_A^1\, q^*(\theta_A^2,\theta_B^2)\Big)\, \nu_A(\theta_B^2)$$
$$=\frac{2}{3}\nu_A(\theta_B^1) - \frac{2}{3}(1-\nu_A(\theta_B^1)).$$

*Thus, (1) if $\nu_A(\theta_B^1) < \mu_A(\theta_B^1) = \frac{1}{2}$, the shortest path from $\theta_0$ to $\theta_A^1$ is $(\theta_0, \theta_A^2, \theta_A^1)$, and (2) if $\nu_A(\theta_B^1) > \mu_A(\theta_B^1) = \frac{1}{2}$, the shortest path from $\theta_0$ to $\theta_A^1$ is $(\theta_0, \theta_A^1)$. It follows that for any $\epsilon > 0$, $q^*$ does not satisfy the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$. By Theorem 1, $S^B > S^R(\epsilon)$, $\forall \epsilon > 0$.*

Example 2 presents an environment in which there would be loss of revenue for the designer as long as the designer has some local uncertainty about agents' beliefs, regardless of how minimal the local uncertainty may be. Upon careful examination, we see that in the network $G(q^*, \mu_A)$, there are multiple shortest paths from $\theta_0$ to $\theta_A^1$ with the same length of 0. Thus, when we consider nearly beliefs, for any $\epsilon > 0$, the shortest-tree property with respect to the uncertainty set $U^\epsilon$ is not satisfied for the decision rule $q^*$.

Fix a decision rule $q$ that is implementable with respect to $\mu$. We say that the shortest path from the dummy type $\theta_0$ to type $\theta_i$ in the network $G(q, \mu_i)$ is unique if there is a unique shortest path $SP(q, \mu_i, \theta_i)$, that is,

$$d(q,\mu_i,\theta_i) = \sum_{(\theta_i',\theta_i'') \in SP(q,\mu_i,\theta_i)} w^{\mu_i}(\theta_i'',\theta_i') < \sum_{(\theta_i',\theta_i'') \in P(q,\mu_i,\theta_i)} w^{\mu_i}(\theta_i'',\theta_i')$$

for any path $P(q, \mu_i, \theta_i) \neq SP(q, \mu_i, \theta_i)$.

**Theorem 2.** *If there exists some optimal decision rule $q^*$ from (BIC) such that for each agent $i$ and each type $\theta_i$, the shortest path from the dummy type $\theta_0$ to type $\theta_i$ in the network $G(q^*, \mu_i)$ is unique, then there exists $\bar{\epsilon} > 0$ such that $S^B = S^R(\epsilon)$ for any $\epsilon \leq \bar{\epsilon}$.*

*Proof.* Fix such a decision rule $q^*$. For each agent $i$ and each type $\theta_i$, we have

$$\begin{aligned} d(q^*,\mu_i,\theta_i) &= \sum_{(\theta_i',\theta_i'') \in SP(q^*,\mu_i,\theta_i)} w^{\mu_i}(\theta_i'',\theta_i') \\ &= \sum_{(\theta_i',\theta_i'') \in SP(q^*,\mu_i,\theta_i)} \Bigg( \sum_{\theta_{-i} \in \Theta_{-i}} \Big( q^*(\theta_i'',\theta_{-i}) \cdot \theta_i'' - q^*(\theta_i',\theta_{-i}) \cdot \theta_i'' \Big)\, \mu_i(\theta_{-i}) \Bigg) \\ &< \sum_{(\theta_i',\theta_i'') \in P} \Bigg( \sum_{\theta_{-i} \in \Theta_{-i}} \Big( q^*(\theta_i'',\theta_{-i}) \cdot \theta_i'' - q^*(\theta_i',\theta_{-i}) \cdot \theta_i'' \Big)\, \mu_i(\theta_{-i}) \Bigg) \end{aligned}$$

$$= \sum_{(\theta_i',\theta_i'')\in P} w^{\mu_i}(\theta_i'',\theta_i').$$

for any path $P$ from the dummy type $\theta_0$ to $\theta_i$ other than $SP(q^*,\mu_i,\theta_i)$. Since the third line holds as an inequality, for any $\nu_i$ sufficiently close to $\mu_i$, $SP(q^*,\mu_i,\theta_i)$ remains the shortest path from the dummy type $\theta_0$ to $\theta_i$ in the network $G(q^*,\nu_i)$, since

$$\sum_{(\theta_i',\theta_i'')\in SP(q^*,\mu_i,\theta_i)} w^{\nu_i}(\theta_i'',\theta_i') < \sum_{(\theta_i',\theta_i'')\in P} w^{\nu_i}(\theta_i'',\theta_i')$$

for any path $P$ from the dummy type $\theta_0$ to $\theta_i$ other than $SP(q^*,\mu_i,\theta_i)$. Therefore, there exists $\bar{\epsilon} > 0$ such that $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$ as long as $\epsilon \leq \bar{\epsilon}$. By Theorem 1, $S^B = S^R(\epsilon)$ for any $\epsilon \leq \bar{\epsilon}$. □

## 5.2 $\lim_{\epsilon\to 0} S^R(\epsilon) = S^B$

In this section, we establish another foundation for the Bayesian approach. Here, the requirement is weaker than that in Section 5.1. Rather than requiring a threshold $\bar{\epsilon} > 0$ such that $S^B = S^R(\epsilon)$ for any $\epsilon \leq \bar{\epsilon}$, we require that the revenue loss due to robustness be vanishingly small as $\epsilon$ approaches 0. In what follows, we show that this holds very generally, as long as a mild slater condition is satisfied.

Let $\mathcal{M}$ the collection of all mechanisms that are Bayesian incentive compatible with respect to $\mu$. We assume that the Bayesian mechanism design problem satisfies the following Slater condition.[12]

**Assumption 1.** *There exists a mechanism $(q,t)$ such that for each agent $i \in \mathcal{I}$, each type $\theta_i \in \Theta_i$, and each $\theta_i' \in \{\Theta_i \setminus \{\theta_i\}\} \cup \{\theta_0\}$,*

$$\sum_{\theta_{-i}\in\Theta_{-i}} \Big(q(\theta_i,\theta_{-i})\cdot\theta_i - t_i(\theta_i,\theta_{-i})\Big)\,\mu_i(\theta_{-i}) > \sum_{\theta_{-i}\in\Theta_{-i}} \Big(q(\theta_i',\theta_{-i})\cdot\theta_i - t_i(\theta_i',\theta_{-i})\Big)\,\mu_i(\theta_{-i}).$$

**Theorem 3.** *Under Assumption 1, $S^R(\epsilon)$ is continuous at the point $\epsilon = 0$, that is,*

$$\lim_{\epsilon\to 0} S^R(\epsilon) = S^B.$$

[12] This is a mild assumption. For instance, in the canonical one-dimensional mechanism design setting, Escude and Sinander (2020) show that (weak) strategy-proofness can be made strict by an arbitrarily small modification.

*Proof.* Theorem 3 follows from continuity arguments. Let $(q^*, t^*)$ be an optimal Bayesian mechanism with respect to $\mu$. Under Assumption 1, $\mathcal{M}$ has non-empty interior. Since $\mathcal{M}$ is convex and closed, for any $\eta > 0$, there exists $(q^{**}, t^{**}) \in \text{ int } \mathcal{M}$ that is sufficiently close to $(q^*, t^*)$ and hence achieves an expected revenue larger than $S^B - \eta$. Since $(q^{**}, t^{**}) \in \text{ int } \mathcal{M}$, there exists $\bar{\epsilon} > 0$ such that as long as $\epsilon \leq \bar{\epsilon}$, $(q^{**}, t^{**})$ is robust incentive compatible with respect to the uncertainty set $U^\epsilon$. It follows that

$$\lim_{\epsilon \to 0} S^R(\epsilon) \geq S^B - \eta.$$

Since $\eta$ is arbitrary, we have

$$\lim_{\epsilon \to 0} S^R(\epsilon) = S^B.$$

□

# 6 Conclusion

This paper presents a systematic analysis of the robust optimization approach to mechanism design within a general social choice environment with quasi-linear preferences and private values. This approach is particularly suited for designing mechanisms when the designer faces local uncertainty about agents' beliefs, which is a common occurrence. We argue that this approach offers two key advantages: compared to the Bayesian approach, it safeguards against potential misspecification of agents' beliefs, and compared to the dominant strategy approach, it typically yields higher expected revenue. By modeling the limit case in which $\epsilon$ tends towards zero, we use the robust optimization approach to rationalize the Bayesian approach.

# A Proof of Proposition 5

In what follows, we show that $q^*$ satisfies the uniform shortest-path tree property with respect to the uncertainty set $U^\epsilon$ if and only if $0 \leq \epsilon \leq \frac{1}{6}$.

For trader $A$, the belief derived from the benchmark prior $\mu$ is $\mu_A(\theta_B^1) = \frac{2}{3}$ and $\mu_A(\theta_B^2) = \frac{1}{3}$. For any $\epsilon \leq \frac{1}{3}$, the uncertainty set $U_A^\epsilon$ has two extreme points:

$$\nu_A(\theta_B^1) = \frac{2}{3} + \epsilon, \quad \nu_A(\theta_B^2) = \frac{1}{3} - \epsilon,$$

$$\nu'_A(\theta_B^1) = \frac{2}{3} - \epsilon, \quad \nu'_A(\theta_B^2) = \frac{1}{3} + \epsilon.$$

*For the network* $G(q, \nu_A)$*:*

The length of the path $(\theta_0, \theta_A^1)$ is 0 and the length of the path $(\theta_0, \theta_A^2, \theta_A^1)$ is

$$\frac{2}{3}(\frac{2}{3} + \epsilon) - \frac{2}{3}(\frac{1}{3} - \epsilon) > 0.$$

Thus, the shortest path from $\theta_0$ to $\theta_A^1$ is $(\theta_0, \theta_A^1)$. The length of the path $(\theta_0, \theta_A^2)$ is

$$\frac{2}{3}(\frac{2}{3} + \epsilon) - \frac{2}{3}(\frac{1}{3} - \epsilon)$$

and the length of the path $(\theta_0, \theta_A^1, \theta_A^2)$ is

$$\frac{4}{3}(\frac{2}{3} + \epsilon).$$

Thus, the shortest path from $\theta_0$ to $\theta_A^2$ is $(\theta_0, \theta_A^2)$.

*For the network* $G(q, \nu'_A)$*:*

The length of the path $(\theta_0, \theta_A^2)$ is

$$\frac{2}{3}(\frac{2}{3} - \epsilon) - \frac{2}{3}(\frac{1}{3} + \epsilon)$$

and the length of the path $(\theta_0, \theta_A^1, \theta_A^2)$ is

$$\frac{4}{3}(\frac{2}{3} - \epsilon).$$

Thus, the shortest path from $\theta_0$ to $\theta_A^2$ is $(\theta_0, \theta_A^2)$. The length of the path $(\theta_0, \theta_A^1)$ is 0 and the length of the path $(\theta_0, \theta_A^2, \theta_A^1)$ is

$$\frac{2}{3}(\frac{2}{3} - \epsilon) - \frac{2}{3}(\frac{1}{3} + \epsilon).$$

Thus, the shortest path from $\theta_0$ to $\theta_A^1$ is $(\theta_0, \theta_A^1)$ if and only if $\epsilon \leq \frac{1}{6}$.

Comparing the two networks $G(q, \nu_A)$ and $G(q, \nu'_A)$, we conclude that there is the same shortest-path tree with respect to the uncertainty set $U^\epsilon$ for trader $A$ if and only if $\epsilon \leq \frac{1}{6}$. Similar calculations for trader $B$ show that there is the same shortest-path tree with respect to the uncertainty set $U^\epsilon$ for trader $B$ if $\epsilon \leq \frac{1}{6}$. Proposition 5 follows from the above calculations and Theorem 1.

# References


Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1993): *Network Flows: Theory, Algorithms and Applications*. Prentice Hall, New Jersey.

Ben-Tal, A., and A. Nemirovski (2002): "Robust Optimization - Methodology and Applications," *Mathematical Programming*, 92(3), 453–480.

Bergemann, D., and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73(6), 1771–1813.

Bergemann, D., and K. Schlag (2011): "Robust Monopoly Pricing," *Journal of Economic Theory*, 146, 2527–2543.

Bertsimas, D., V. Gupta, and N. Kallus (2018): "Data-Driven Robust Optimization," *Mathematical programming*, 167, 235–292.

Beyer, H.-G., and B. Sendhoff (2007): "Robust Optimization - A Comprehensive Survey," *Computer methods in applied mechanics and engineering*, 196(33), 3190–3218.

Börgers, T. (2015): *An Introduction to the Theory of Mechanism Design*. Oxford University Press.

Brooks, B., and S. Du (2021): "Optimal Auction Design with Common Values: An Informationally-Robust Approach," *Econometrica*, 89(3), 1313–1360.

Carroll, G. (2017): "Robustness and Separation in Multidimensional Screening," *Econometrica*, 85(2), 453–488.

Carroll, G., and D. Meng (2016): "Locally Robust Contracts for Moral Hazard," *Journal of Mathematical Economics*, 62, 36–51.

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," *Review of Economic Studies*, 74(2), 447–476.

Cramton, P., R. Gibbons, and P. Klemperer (1987): "Dissolving a Partnership Efficiently," *Econometrica*, 55(3), 615–632.

Cremer, J., and R. P. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56(6), 1247–1257.

Cui, T., Y.-J. Chen, and Z.-J. M. Shen (2009): "Optimal Auction Design: Uncertainty, Robustness, and Revenue Maximization," working paper.

Du, S. (2018): "Robust Mechanisms under Common Valuation," *Econometrica*, 86(5), 1569–1588.

Escude, M., and L. Sinander (2020): "Strictly Strategy-proof Auctions," *Mathematical Social Sciences*, 107, 13–16.

Gershkov, A., J. Goeree, A. Kushnir, B. Moldovanu, and X. Shi (2013): "On the Equivalence of Bayesian and Dominant Strategy Implementation," *Econometrica*, 81(1), 197–220.

Heydenreich, B., R. Müller, M. Uetz, and R. Vohra (2009): "Characterization of Revenue Equivalence," *Econometrica*, 77(1), 307–316.

Hiriart-Urruty, J.-B., and C. Lemaréchal (2004): *Fundamentals of Convex Analysis*. Springer Science & Business Media.

Jehiel, P., M. Meyer-ter Vehn, and B. Moldovanu (2012): "Locally Robust Implementation and its Limits," *Journal of Economic Theory*, 147, 2439–2452.

Li, J., and P. Dworczak (2024): "Are Simple Mechanisms Optimal when Agents are Unsophisticated?," working paper.

Loertscher, S., and L. M. Marx (2020): "A Dominant-Strategy Asset Market Mechanism," *Games and Economic Behavior*, 120, 1–15.

Lopomo, G., L. Rigotti, and C. Shannon (2020): "Uncertainty in Mechanism Design," working paper.

——— (2022): "Uncertainty and Robustness of Surplus Extraction," *Journal of Economic Theory*, 199, 105088.

Lu, H., and J. Robert (2001): "Optimal Trading Mechanisms with Ex Ante Unidentified Traders," *Journal of Economic Theory*, 97(1), 50–80.

Madarász, K., and A. Prat (2017): "Sellers with Misspecified Models," *Review of Economic Studies*, 84, 790–815.

Mailath, G., and A. Postlewaite (1990): "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies*, 57(3), 351–367.

Malakhov, A., and R. V. Vohra (2009): "An Optimal Auction for Capacity Constrained Bidders: A Network Perspective," *Economic Theory*, 39(1), 113–128.

Manelli, A. M., and D. R. Vincent (2010): "Bayesian and Dominant-Strategy Implementation in the Independent Private-Values Model," *Econometrica*, 78(6), 1905–1938.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

Myerson, R., and M. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29(2), 265–281.

Ollár, M., and A. Penta (2017): "Full Implementation and Belief Restrictions," *American Economic Review*, 107(8), 2243–2277.

——— (2023): "A Network Solution to Robust Implementation: The Case of Identical but Unknown Distributions," *Review of Economic Studies*, 90(5), 2517–2554.

Pham, H., and T. Yamashita (2024): "Auction Design with Heterogeneous Priors," *Games and Economic Behavior*, 145, 413–425.

Rochet, J.-C. (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context," *Journal of Mathematical Economics*, 16(2), 191–200.

Vohra, R. (2011): *Mechanism Design: A Linear Programming Approach.* Cambridge University Press.

Yamashita, T., and S. Zhu (2022): "On the Foundations of Ex Post Incentive-Compatible Mechanisms," *American Economic Journal: Microeconomics*, 14(4), 494–514.