WORKING PAPERS

N° 1260

March 2023

"Auction Design with Heterogeneous Priors"

Hien Pham and Takuro Yamashita

Toulouse School of Economics

Auction Design with Heterogeneous Priors*

Hien Pham† Takuro Yamashita‡

March 22, 2023

Abstract

We consider an auction design problem with private values, where the seller and bidders may enjoy heterogeneous priors about their (possibly correlated) valuations. Each bidder forms an (interim) belief about the others based on his own prior updated by observing his own value. If the seller faces uncertainty about the bidders' priors, even if he knows that the bidders' priors are within any given distance from his, he may find it worst-case optimal to propose a dominant-strategy auction mechanism.

*We are grateful to Gabriel Carroll, Jacques Crémer, Ulrich Hege, Christian Hellwig, Antonio Penta, and participants at several workshops, seminars, and conferences. We acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 714693), and ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

†Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. hien.pham@tse-fr.eu

‡Osaka School of International Public Policy, Osaka University, Machikaneyama 1-31, Toyonaka, Osaka, 560-0043, Japan. yamashitat@osipp.osaka-u.ac.jp

1 Introduction

The common knowledge assumptions have been challenged by many papers in the literature on robust mechanism design (see our detailed discussion in the related literature section). In an influential work, Chung and Ely (2007) consider an auction environment where the seller has little idea about each bidder's belief about the other bidders' valuations. They show that, for some specification of the bidders' beliefs (formally identified by a type space), a dominant-strategy auction mechanism is revenue-maximizing among all Bayesian incentive compatible auction mechanisms, even if the seller knows that that type space governs the bidders' beliefs (*Baysian foundation* for a dominant-strategy mechanism). As a consequence, in case the seller does not know which type space governs their beliefs, a dominant-strategy auction mechanism is max-min optimal (*Maximin foundation*).

The interim belief that each bidder of each type must have in this critical type space is special, and seems very different from what any common-prior type space would imply. Indeed, Chung and Ely (2007) provide a counterexample such that a dominant-strategy auction mechanism cannot be Bayesian-founded if the bidders' type space must be one of the common-prior type spaces.

A natural question is "how far" this Chung-Ely's type space is relative to those given by some common prior. To investigate this question, we examine the class of types spaces which are induced by *(ϵ-) heterogeneous priors*. Namely, each player (seller and each bidder) possesses a prior distribution about the value distribution before their values being drawn, which can be ε-different from each other (in the metric similar to the one considered by Madarász and Prat (2017) and Carroll (2017)). Then, each bidder's value is drawn, making him Bayesian update his own prior conditional on his own value. Clearly, with $\varepsilon = 0$, the model reduces to the standard common-prior case, and hence, no foundation. With a large enough ε, it is natural to think that the critical type space of Chung-Ely can be captured, and hence a foundation exists. We show that, in fact, the critical type space of Chung-Ely can be represented by a type space induced by ϵ-heterogenous priors, *for any* $\varepsilon > 0$, no matter how small it is. Therefore, with any $\varepsilon > 0$, the dominant-strategy auction mechanism is Bayesian (and hence Maximin) founded.

The basic intuition is that, even if a bidder's prior is close to the others' (in particular, to the seller's), it does not mean that their "interim beliefs" are close to each other. In fact, they can be so flexible that any small (but positive) heterogeneity in their priors can result in very different interim beliefs.

Although the original result of Chung and Ely (2007) suggests that the dominant-strategy approach would be reasonable in case the seller has *very little* idea about the bidders' information (for example, when there have not been similar items auctioned), it is sometimes informally argued that, if rich data is available about past similar auctions, it might be more difficult to justify the dominant-strategy approach, as both the seller and bidders would have a more precise idea about the true value distribution. In practice, the players typically have *some* information about past similar auctions, though they never have an *exact* common prior. In this sense, it is important to investigate the "boundary" of Chung-Ely's argument: With which class of type spaces (related to which information of the bidders about past similar auctions) the dominant-strategy approach has a Chung-Ely foundation? The result of our paper contributes to a better understanding of this question by examining (possibly small) heterogeneity in the players' priors.

1.1 Related literature

This paper contributes to the growing literature on robust mechanism design (see, for example, Bergemann and Morris (2005), Chung and Ely (2007), Chen and Li (2018), and Yamashita and Zhu (2022) as the most relevant ones to this paper). These papers consider the situation where the agents' beliefs can be arbitrarily different from each other (and from the principal's, if the principal has a prior). For example, as aforementioned, Chung and Ely (2007) identifies a type space with heterogeneous priors with which one of the optimal Bayesian mechanisms is a dominant-strategy mechanism. Thus, if the seller has little idea about the bidders' beliefs, then the worst-case-minded seller has a justification to offer a dominant-strategy mechanism. See Chen and Li (2018) for its generalization to non-auction environments. We show that, even if the seller has a much better idea about the bidders' beliefs in that their priors are arbitrarily close to each other and also to the seller's (and that being their common knowledge), essentially the same conclusion is obtained. In this sense, our result strengthens that of Chung and Ely (2007).

Our notion of prior perturbations is related to the (various) notions of "local robustness" in the literature. For example, in Lopomo, Rigotti, and Shannon (2021) where each agent's type is associated with a set of "fully overlapping"[1] interim beliefs, a mechanism is robust if it is implementable for every possible interim belief. They find that robustness is hard to achieve even when this set is arbitrarily small. As another example, Ollár and Penta (2017) propose a general form of restrictions directly on the agents' interim beliefs, and show that, when the set of possible interim

[1]Roughly speaking, this "fully overlapping" requirement means that nearby types share a sufficiently rich set of beliefs. A focal special case is when the set is an arbitrarily small neighborhood around a fixed belief. See also Lopomo, Rigotti, and Shannon (2022) where they derive the necessary and sufficient conditions for full extraction in this setting.

beliefs is small in an appropriate sense, much more permissive results are possible. Our result shows that the *ex ante* belief restriction does not imply their *interim* restriction, and hence they lead to very different results. In this sense, our notion of uncertainty may be interpreted as "ex-ante-local" uncertainty. Jehiel, Meyer-ter Vehn, and Moldovanu (2012) consider a related notion of local uncertainty in terms of interim beliefs, but in a generic multi-dimensional interdependent-value environment. They show that, if the principal's goal is to implement some belief-invariant social choice function, then the same kind of an impossibility result is obtained as in Jehiel, ter Vehn, Moldovanu, and Zame (2006) (where the latter paper considers ex post implementation, and in this sense allows for global robustness). Our environment is with private values, and the seller's goal is revenue maximization rather than a social choice function implementation.[2]

In a single-agent environment, Madarász and Prat (2017) consider a situation where the principal is aware that the true distribution of the agent's type can be ε-different from what the seller has in mind. Carroll (2017) generalizes their notion of ε-closeness in the context of a (single-agent) multi-dimensional screening problem.[3] As far as we are aware, ours is the first paper that generalizes their notions of closeness to a multi-agent environment. Importantly, with multiple agents, it is not only the principal who is uncertain about the true distribution, but also the agents enjoy uncertainties about the true distributions and the others' beliefs. On the other hand, relative to Madarász and Prat (2017) and Carroll (2017), we focus on

[2]Hence, in principle, the seller might find it optimal to use a mechanism that induces a highly belief-dependent outcome. Put differently, the set of feasible mechanisms in our case is larger than those that implement a belief-invariant social choice function.

[3]See also Bergemann and Schlag (2011). Carroll and Meng (2016) considers the local robustness in a single-agent moral-hazard environment.

a single-good private-value auction, with which the agents' payoff structures satisfy the single-crossing conditions.

There has been some work on mechanism design with heterogeneous priors. For example, in Eliaz and Spiegler (2008), a consumer assigns excessive weights to the states of nature associated with their large gains from trade. They find that non-common priors can be necessary for price discrimination. Grubb (2009) studies a situation where a consumer assigns wrong weights to their possible valuations (narrowly concentrates around the mean), relative to the seller's prior. He mainly focuses on characterizing the optimal contract under complete information. Our paper introduces heterogeneous priors in the auction context (i.e., with multiple agents rather than a single representative agent).

2 Auction environment

A seller wants to sell an indivisible good. There are N risk-neutral bidders with private values. Each bidder $i \in \{1, \ldots, N\}$ knows his own valuation $v_i \in \mathbb{R}$. An allocation is denoted by $(q, p) = (q_i, p_i)_{i=1}^{N}$, where $q_i \in [0, 1]$ denotes the probability that bidder i obtains the good, and $p_i \in \mathbb{R}$ denotes his payment to the seller. An allocation is feasible if $\sum_i q_i \leq 1$. Given (q_i, p_i), i's payoff is given by $v_i q_i - p_i$.

The players (the seller and the bidders) enjoy *heterogeneous priors* for the distribution of the bidders' values v. Specifically, let $g \in \Delta(\mathbb{R}^N)$ be the seller's prior, which has a finite support represented by $\{\gamma, 2\gamma, \ldots, K\gamma\}^N$ (following Chung and Ely (2007)) for some $K \in \mathbb{N}$ and $\gamma > 0$ for notational simplicity. Throughout the paper, we assume that g satisfies the single-crossing virtual value condition (Chung and Ely (2007)). For each $i \neq j$ and v, let $\gamma_i(v)$ be i's virtual valuation:

$$\gamma_i(v) = v_i - \gamma \frac{1 - G_i(v)}{g(v)},$$

where $G_i(v) = \sum_{v'_i \leq v_i} g(v'_i, v_{-i})$.

Assumption 1. *For each $i \neq j$, and each v_i, v'_i, v_{-i} with $v'_i > v_i$:*

$$\gamma_i(v_i, v_{-i}) \geq 0 \quad \Rightarrow \quad \gamma_i(v'_i, v_{-i}) > 0$$
$$\gamma_i(v_i, v_{-i}) \geq \gamma_j(v_i, v_{-i}) \quad \Rightarrow \quad \gamma_i(v'_i, v_{-i}) > \gamma_j(v'_i, v_{-i}).$$

As shown in Chung and Ely (2007), it is satisfied if g exhibits affiliation and monotone hazard-rates. In this sense, it may be considered a mild assumption.

As opposed to the standard exact-common-prior model where not only the seller but every bidder i believes g (and that itself being common knowledge), we allow the possibility that they enjoy heterogeneous priors: For each i, let $h_i \in \Delta(\mathbb{R}^N)$ be bidder i's prior, which again has a finite support for simplicity (but potentially with a different support from g and from h_j, $j \neq i$). Note that bidder i knows his own value v_i at the time he plays an auction mechanism. That his prior is h_i implies that his belief about the others' values is based on h_i conditional on his v_i.

We assume that the seller has limited knowledge as to "how distant" each bidder i's h_i could be from the seller's prior g. This distance may be interpreted as the level of the seller's confidence in his own information.[4]

Our notion of distance is based on Madarász and Prat (2017) and Carroll (2017):

[4]This interpretation implies a related but different question: what if the seller's prior g is different from the true value distribution? For now, we assume that the seller is confident in his own g as the true value distribution, but we study the case where the seller fears the possibility that g is wrong. See Section 6.

Definition 1. *Two distributions μ and $\hat{\mu}$ are ε-close to each other if $V = supp(\mu)$ and $\hat{V} = supp(\hat{\mu})$ can be partitioned into disjoint measurable sets $\{V^1, ..., V^r\}$ and $\{\hat{V}^1, ..., \hat{V}^r\}$ respectively such that, for each $k \in \{1, ..., r\}$:*

1. $\mu(V^k) = \hat{\mu}(\hat{V}^k)$, *and*
2. $d(v, \hat{v}) \leq \varepsilon$ *for any* $(v, \hat{v}) \in V^k \times \hat{V}^k$,

where $d(v, \hat{v})$ represents the Euclidean distance between v and $\hat{v}$.

A collection of distributions $\{\mu_1, \ldots, \mu_K\}$ is ε-close to each other if any pair μ_i, μ_j are ε-close to each other as above.

Example 1. *We illustrate the closeness of two distributions in the following example with $N = 2$. Let g be the distribution represented as follows:*

$g(v_1, v_2)$	$v_2 = 1$	$v_2 = 2$
$v_1 = 1$	$\frac{1}{3}$	$\frac{1}{6}$
$v_1 = 2$	$\frac{1}{6}$	$\frac{1}{3}$

Table 1: Distribution g

and let f be represented as follows:

$f(v_1, v_2)$	$v_2 = 1 - \varepsilon$	$v_2 = 1$	$v_2 = 2 - \varepsilon$	$v_2 = 2$
$v_1 = 1 - \varepsilon$	$\frac{1}{3}$			
$v_1 = 1$			$\frac{1}{6}$	
$v_1 = 2 - \varepsilon$		$\frac{1}{6}$		
$v_1 = 2$				$\frac{1}{3}$

Table 2: Distribution f

Then, according to the definition above, f and g are $(\varepsilon\sqrt{2})$-close to each other. Figures 1 and 2 illustrate f and g in the (v_1, v_2)-space.

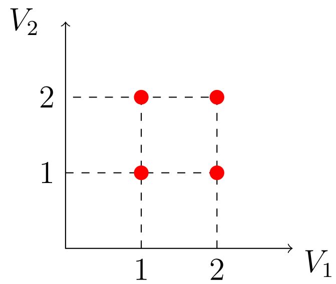

Figure 1: Distribution g

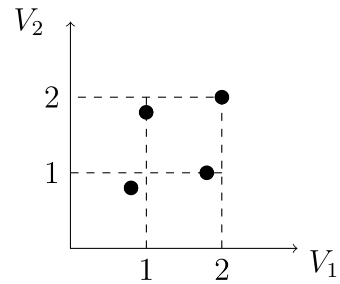

Figure 2: Distribution f

The seller believes that $(g, h_1, \ldots, h_N)$ are ε-close to each other for some given $\varepsilon > 0$, and that any such combination of h_i's is possible. This uncertainty makes the seller cautious in designing an auction mechanism.

3 Auction mechanism

Given the concern about the above prior heterogeneity, the seller can design a "robust" mechanism in a certain sense. One of the possible approaches is to design a *dominant-strategy* auction mechanism, where each bidder has a dominant action given each v_i regardless of the other bidders' behavior. Such a mechanism can guarantee some level of expected revenue regardless of each bidder's belief about the opponents' values and their (higher-order) beliefs; in particular, regardless of each i's prior h_i.

Another possibility is to try to extract each bidder's information (about each i's h_i, for example) in order to design a more profitable auction mechanism. Indeed, in the standard exact common-prior environment where $g = h_i$ for all i is common knowledge, if $g(= h_i)$ satisfies a certain correlation structure, the seller can extract the entire surplus (Cremer-McLean), while the optimal dominant-strategy mechanism leaves a non-negligible rent to the winning bidder. Even if h_i can be different from g, if the seller knows that they cannot be too far from each other, it may be natural to expect that better mechanisms than dominant-strategy mechanisms exist.

3.1 Notation

An auction mechanism is represented by $(M, q, p) = (M_i, q_i, p_i)_{i=1}^N$, where: each M_i is a set, $M = \prod_{i=1}^N M_i$, $q_i : M \to [0, 1]$ with $\sum_i q_i(m) \le 1$ for all $m \in M$, and $p_i : M \to \mathbb{R}$. An interpretation is that, given mechanism (M, q, p), each bidder is asked to simultaneously choose any $m_i \in M_i$; and given a chosen vector $m = (m_1, \ldots, m_N) \in M$, allocation $(q_i(m), p_i(m))_{i=1}^N$ is executed. A feasible mechanism must contain some element $\phi_i \in M_i$ for each i such that $q_i(\phi_i, m_{-i}) = p_i(\phi_i, m_{-i}) = 0$ for any $m_{-i} \in M_{-i}$, representing the idea of i's individual rationality requirement.

3.2 Dominant-strategy auction mechanism

We first introduce *dominant-strategy auction mechanisms.*

Definition 2. *Mechanism (M, q, p) admits a dominant-strategy equilibrium if there exists $\sigma_i(v_i)$ for each $i, v_i \in \mathbb{R}$ such that, for each m_i, m_{-i}:*

$$\begin{aligned} v_i q_i(\sigma_i(v_i), m_{-i}) - p_i(\sigma_i(v_i), m_{-i}) &\ge v_i q_i(m_i, m_{-i}) - p_i(m_i, m_{-i}) \\ v_i q_i(\sigma_i(v_i), m_{-i}) - p_i(\sigma_i(v_i), m_{-i}) &\ge 0. \end{aligned}$$

Mechanism Γ guarantees expected revenue R in dominant strategy if Γ admits a *dominant-strategy equilibrium* $\sigma = (\sigma_i)_{i=1}^N$ *such that*

$$\sum_v [\sum_i p_i(\sigma(v))] g(v) \geq R.$$

Let R^D denote the best revenue guarantee in dominant strategy. That is, for any $R < R^D$, there is a mechanism which guarantees R in dominant strategy.

3.3 Bayesian auction mechanism

In order to define the other standard concept of Bayesian equilibrium, we need further information about the bidders' higher-order beliefs (such as what each bidder believes about the others' values, and about the others' beliefs about it, etc.). In this paper, we consider the simplest possible alternative: Each bidder i believes h_i as his first-order belief, and that fact itself is common knowledge (i.e., trivial higher-order beliefs).[5]

Given that $(h_i)_{i=1}^N$ is common knowledge among the bidders, a Bayesian equilibrium in a mechanism is naturally defined as follows.

Definition 3. *Mechanism* Γ *admits a Bayesian equilibrium given* $(h_i)_{i=1}^N$ *if there exists* $\sigma_i(v_i)$ *for each* $i, v_i \in \mathbb{R}$ *such that, for each* m_i*:*

$$\begin{aligned} & \sum_{v_{-i}} [v_i q_i(\sigma_i(v_i), \sigma_{-i}(v_{-i})) - p_i(\sigma_i(v_i), \sigma_{-i}(v_{-i}))] h_i(v_i, v_{-i}) \\ \geq & \sum_{v_{-i}} [v_i q_i(m_i, \sigma_{-i}(v_{-i})) - p_i(m_i, \sigma_{-i}(v_{-i}))] h_i(v_i, v_{-i}). \end{aligned}$$

[5]See Appendix (Section A) for the formal description of the type space considered here. Our modelling choice may be justified as follows. First, even if one prefers other specifications, they would probably include this common-knowledge possibility as one of the possible situations; Second, as a related point, our approach would make the departure from the standard exact-common-knowledge model minimal. Given that our result is basically a negative result, this minimalistic choice makes the conclusion strongest.

Mechanism Γ *guarantees expected revenue* R_ε *in Bayesian equlibrium if, for any* $(h_i)_{i=1}^N$ *such that* $(g, (h_i)_{i=1}^N)$ *are* ε*-close to each other,* Γ *admits a Bayesian equilibrium* $\sigma = (\sigma_i)_{i=1}^N$ *given* $(h_i)_{i=1}^N$ *such that*

$$\sum_v [\sum_i p_i(\sigma(v))]g(v) \geq R.$$

Let $R_\varepsilon^\star$ denote the best revenue guarantee in Bayesian equilibrium. That is, for any $R_\varepsilon < R_\varepsilon^\star$, there is a mechanism which guarantees R_ε in Bayesian equilibrium.

Obviously, the best revenue guarantee in dominant strategy is weakly lower than that in Bayesian equilibrium: For any ε,

$$R_\varepsilon^\star \geq R^D.$$

Recall that, in case g exhibits certain correlation (as specified in Crémer and McLean (1988)) and $\varepsilon = 0$, the expected revenue in Bayesian equilibrium is very different from that in dominant strategies (i.e., $R_0^\star > R^D$). On the contrary, we show that, as long as ε is strictly positive, no matter how small it is, the guaranteed revenue in Bayesian equilibrium coincides with that in dominant strategies (i.e., $R_\varepsilon^\star = R^D$).

We prove this claim in Section 5, followed by a motivating example in Section 4, explaining why the problem with $\varepsilon > 0$ can be very different from that with $\varepsilon = 0$.

4 Motivating example

We employ Example 1 to illustrate the seller's revenue loss if he adopts the optimal mechanism *without* taking into account the possibility of prior heterogeneity. More precisely, imagine that the seller wrongly assumes that g is the common prior, while each bidder i actually has a different prior $h_i \neq g$. We will show that the seller's revenue loss does not vanish even when g and each h_i get closer in the sense of our distance.

Assume that the seller's benchmark distribution g is as illustrated in Table 1. If the seller believes that g is the common prior, then as in Crémer and McLean (1988), the optimal mechanism is a combination of a second-price auction (SPA) and side-bets, which extracts the full surplus as his expected revenue ($\frac{5}{3}$). The following table corresponds to one such mechanism (it only shows bidder 1's allocation; bidder 2's is symmetric), where "NP" stands for non-participation:

$(q_1(v), t_1(v))$	NP	$v_2 = 1$	$v_2 = 2$
NP	$(0, 0)$	$(0, 0)$	$(0, 0)$
$v_1 = 1$	$(1, 0)$	$(\frac{1}{2}, \frac{1}{2} - \frac{1}{3})$	$(0; \frac{2}{3})$
$v_1 = 2$	$(1, 0)$	$(1, 1 - \frac{1}{3})$	$(\frac{1}{2}, 1 + \frac{2}{3})$

Table 3: Outcomes from a SPA and side-bets

where the red parts in the transfers come from the side-bets. Each bidder's expected payment is $\frac{1}{3}\left(\frac{1}{2} - \frac{1}{3} + 1 + \frac{2}{3}\right) + \frac{1}{6}\left(1 - \frac{1}{3} + \frac{2}{3}\right) = \frac{5}{6}$, and therefore, the expected revenue is $\frac{5}{3}$, which is exactly the ex-ante total surplus.

Now consider the case where each bidder i's prior h_i is ε-close to but different from g. One might conjecture that, if the above mechanism is appropriately perturbed so that the bidders' participation and incentive constraints are satisfied *with strict inequality* (more specifically, with the strictness in the order of ε), then a similar level of expected revenue may be guaranteed. In particular, as $\varepsilon \to 0$, that guaranteed revenue converges to the full-surplus revenue again.

This conjecture is false. To explain the key idea, suppose that each h_i coincides with f in Table 2, while g is, as assumed by the seller, the true distribution of values. Even though f and g are ε-close to each other as *priors*, they are very different in terms of their induced conditional distributions, that is, each bidder's *interim* belief given his value. Given $f(= h_i)$, bidder i with any v_i essentially *knows* the other bidder's value. Therefore, in the above Crémer-McLean mechanism, truth-telling (or more precisely, reporting the values closest to their true values) is no longer an equilibrium.

For example, bidder i with $v_i = 1$ puts probability 1 on bidder $-i$'s having $v_{-i} = 2 - \varepsilon$, and vice versa. They play a " complete-information" equilibrium where bidder $-i$ bids 2 and *bidder i does not participate in the auction.* Similarly, bidder i with $v_i = 2$, putting probability 1 on $v_{-i} = 2$, does not participate in the auction either. Therefore, as long as g is the true distribution (which only assigns positive probabilities on $v_i \in \{1, 2\}$), no one participates in the auction, yielding 0 revenue.

Note that this property does not depend on the exact value of $\varepsilon > 0$. Therefore, the seller's expected revenue in this mechanism would be far below the first-best surplus.

5 Main result

In this section, we show that $R^D = R^\star_\varepsilon$ for any $\varepsilon > 0$.

Theorem 1. *For any $\varepsilon > 0$, we have:*

$$R^D = R^\star_\varepsilon.$$

The proof is in Appendix B, and proceeds as follows. The key intuition is that, even if $\varepsilon(> 0)$ is arbitrarily small, it is always possible to find a specific prior h_i of each bidder i such that, after Bayesian updating observing i's own value v_i, his "interim belief" about the others' values is very different from the one where i's prior is g (i.e., the case with $\varepsilon = 0$). Moreover, this interim belief structure is such that the seller finds it optimal to offer a dominant-strategy auction mechanism *even if he knows that that h_i is each bidder's prior*. This last property is building on the original work by Chung and Ely (2007), while our more concise proof is building on Chen and Li (2018).

Recall that the original result of Chung and Ely (2007) shows that an auction seller finds it optimal to offer a dominant-strategy auction mechanism to bidders if the seller has very little idea as to the bidders' belief structure, and hence any interim belief structure is deemed possible. Our result suggests that their result is relevant not only when the seller literally has very little idea about the bidders' information, but also when the seller and bidders have close (but heterogeneous) priors.

6 Possible misspecification of g

So far, the seller assumes that his prior g is the true distribution of the bidders' valuations, although he thinks it possible that the bidders' priors are "ε-different" from g. However, if we interpret this ε as the seller's level of confidence in his g, it may also be natural to allow for the seller to worry about the possibility that g is not the true distribution of valuations.

Formally, let f represent the true distribution of v, the bidders' value profile. The seller does not know f, while he thinks that his prior g is a reasonable approximation of f (and each bidder's prior h_i). Based on the idea that $\varepsilon(> 0)$ represents the seller's confidence in his g, we assume that $(f, g, (h_i)_{i=1}^N)$ are ε-close to each other.

To explain the subtlety, consider the optimal dominant-strategy mechanism if g is indeed the true prior (which guarantee R^D). Typically, some incentive compatibility constraints are binding in this mechanism. Thus, if $f(\neq g)$ is the true prior with $\text{supp}(f) \neq \text{supp}(g)$, some bidders may find it strictly optimal to make a type report that is far from his true type.[6]

Nevertheless, we show that an appropriately modified version of the mechanism, which we call a *transfer-reducing mechanism*, guarantees the same level of expected revenue even if $f \neq g$, as ε vanishes. The key of the construction is, by reducing the transfers of the mechanism (by an appropriate amount as a function of ε), the mechanism can now make all the incentive constraints satisfied in a stronger sense, so that even if g and f have (ε-)different supports, each agent finds it dominant to report the value that is closest to his true value in that mechanism. Although the revenue must be smaller, as $\varepsilon \to 0$, this revenue loss vanishes.

Formally, let R^D_ε represent the optimal revenue guarantee in dominant strategy if $(f, g, (h_i)_{i=1}^N)$ are ε-close to each other.

Theorem 2. *$R^D_\varepsilon \to R^D$ as $\varepsilon \to 0$.*

Madarász and Prat (2017) show that, in a general single-agent mechanism design environment, a similar approximation result is possible by their *profit-participation mechanism* even without single-crossing conditions. That is, as the seller's benchmark distribution converges to the true distribution, their optimal expected revenues also converge. Its basic idea is to make the agent "biased in favor of the principal" so that any (even non-local) deviation due to misspecification only increases the principal's payoff. Our proof generalizes their result to a multi-agent environment, but in a single-crossing payoff environment. The single-crossing property seems important for this continuity result with multiple agents. To explain this, it is worth noting that (a naive adaptation of) their profit-participating mechanism may not work in our multi-agent setup. This is because, under that mechanism, each agent might have an

[6] In Appendix C, we observe that such a global deviation under misspecification is the norm rather than the exception.

incentive to deviate globally (i.e., a value far from his true one is reported), which in turn distorts other agents' reporting strategies. Consequently, it is not certain that the vanishing revenue loss is obtained. Our transfer-reducing mechanism prevents such global deviations by ensuring that it is a dominant strategy for agents to report the value closest to their true values.

7 Conclusion

In this paper, we consider the private-value auction setting where the true distribution of bidders' valuations is unknown. The seller and each bidder, however, know its approximation. In this framework, we have shown that the dominant-strategy mechanism secures the seller with the highest revenue guarantee. Besides, if the seller is restricted to using a dominant-strategy mechanism, we have characterized the transfer reducing mechanism that helps the seller to obtain a vanishing loss as the estimates by her and the bidders get close to the truth.

There are several follow-up questions. Firstly, when restricting to dominant-strategy mechanisms, our proof works only if the bidders' payoff functions satisfy the single-crossing condition. Although this property holds for a wide range of mechanism design problems, there are cases where it does not hold, such as multi-unit auctions. In such situations, our proposed mechanism may not work.

Another natural direction is to characterize the optimal robust mechanisms in non-auction environments[7] or with common/interdependent values.[8] We leave these

[7]Chen and Li (2018) generalize the foundation result of Chung and Ely (2007) to some private-value non-auction environments. We conjecture that our approach would work in those environments, establishing the worst-case optimality of dominant-strategy mechanisms.

[8]Yamashita and Zhu (2022) generalize the foundation result of Chung and Ely (2007) to an interdependent-value auction environment. We conjecture that our approach would work in those

potential extensions for future research.

environments. However, they suggest that general interdependent-value models may not admit the same sort of foundation result, and in those cases, it is an open question how the approximate worst-case optimal mechanism would look like.

A Formal description of the type space in Definition 3

The type space we consider in Definition 3 is in the class of the known-own-payoff-type type space (Bergemann and Morris (2005)), denoted by $(T_i, \hat{v}_i, \hat{\beta}_i)_{i=1}^N$. For each i, let (i) $T_i = \text{supp}\{v_i | \exists v_{-i};\ h_i(v_i, v_{-i}) > 0\}$, (ii) $\hat{v}_i : T_i \to \mathbb{R}$ be an identity map (i.e., $\hat{v}_i(t_i) = t_i$ for all $t_i \in T_i$), and (iii) $\hat{\beta}_i : T_i \to \Delta(T_{-i})$ is consistent with h_i in the sense that:

$$\hat{\beta}_i(t_{-i}|t_i) = \frac{h_i(t_i, t_{-i})}{\sum_{t'_{-i}} h_i(t_i, t'_{-i})}.$$

B Proof of Theorem 1

We construct each bidder i's prior, h_i, as follows.

For each i, let $V_i = \{v_i \mid \exists v_{-i},\ g(v_i, v_{-i}) > 0\}$ denote the set of i's possible values in the true distribution g, and denote it by $V_i = \{v_i^1, \ldots, v_i^m, \ldots, v_i^M\}$ so that $v_i^m < v_i^{m+1}$. Define $\hat{V}_i = \{v_i + \varepsilon \mid v_i \in V_i\}$ as the "shifted" version of V_i by ε. Also, we denote $V_{-i} = \{v_{-i}^1, \ldots, v_{-i}^k, \ldots, v_{-i}^K\}$, without any ordering on them (i.e., arbitrary labelling will do). Define $h_i(\cdot)$ so that: for each m, k,

$$h_i(v_i^m, v_{-i}^k) = x_i \frac{\tau_i^\star(v_{-i}^k | v_i^m)}{\tau_i^\star(v_{-i}^1 | v_i^m)}$$

(recall $v_i^m \in V_i$), and

$$h_i(v_i^m + \varepsilon, v_{-i}^k) = g(v_i^m, v_{-i}^k) - h_i(v_i^m, v_{-i}^k)$$

(recall $v_i^m + \varepsilon \in \hat{V}_i$), where

$$\tau_i^\star(v_{-i}|v_i) \equiv \frac{\sum_{\hat{v}_i \geq v_i} g(\hat{v}_i, v_{-i})}{\sum_{v_{-i}} \sum_{\hat{v}_i \geq v_i} g(\hat{v}_i, v_{-i})}, \quad x_i = \min_{k,m} \frac{\tau_i^\star(v_{-i}^1 | v_i^m)}{\tau_i^\star(v_{-i}^k | v_i^m)} g(v_i^m, v_{-i}^k).$$

The following table illustrates our construction:

$h_i(.,.)$	v^1_{-i}	v^2_{-i}	...	v^K_{-i}
v_i^1	x_i	$x_i\frac{\tau_i^\star(v^2_{-i}\vert v_i^1)}{\tau_1^\star(v^1_{-i}\vert v_i^1)}$	...	$x_i\frac{\tau_i^\star(v^K_{-i}\vert v_i^1)}{\tau_1^\star(v^1_{-i}\vert v_i^1)}$
$v_i^1+\varepsilon$	$g(v_i^1,v^1_{-i})-x_i$	$g(v_i^1,v^2_{-i})-x_i\frac{\tau_i^\star(v^2_{-i}\vert v_i^1)}{\tau_i^\star(v^1_{-i}\vert v_i^1)}$	...	$g(v_i^1,v^K_{-i})-x_i\frac{\tau_i^\star(v^K_{-i}\vert v_i^1)}{\tau_1^\star(v^1_{-i}\vert v_i^1)}$
v_i^2	x_i	$x_i\frac{\tau_i^\star(v^2_{-i}\vert v_i^2)}{\tau_i^\star(v^1_{-i}\vert v_i^2)}$	...	$x_i\frac{\tau_i^\star(v^K_{-i}\vert v_i^2)}{\tau_i^\star(v^1_{-i}\vert v_i^2)}$
$v_i^2+\varepsilon$	$g(v_i^2,v^1_{-i})-x_i$	$g(v_i^2,v^2_{-i})-x_i\frac{\tau_i^\star(v^2_{-i}\vert v_i^2)}{\tau_i^\star(v^1_{-i}\vert v_i^2)}$	...	$g(v_i^2,v^K_{-i})-x_i\frac{\tau_1^\star(v^K_{-i}\vert v_i^2)}{\tau_i^\star(v^1_{-i}\vert v_i^2)}$
...	...	...	...	...
v_i^M	x_i	$x_i\frac{\tau_i^\star(v^2_{-i}\vert v_i^M)}{\tau_i^\star(v^1_{-i}\vert v_i^M)}$	...	$x_i\frac{\tau_i^\star(v^K_{-i}\vert v_i^M)}{\tau_i^\star(v^1_{-i}\vert v_i^M)}$
$v_i^M+\varepsilon$	$g(v_i^M,v^1_{-i})-x_i$	$g(v_i^M,v^2_{-i})-x_i\frac{\tau_i^\star(v^2_{-i}\vert v_i^M)}{\tau_i^\star(v^1_{-i}\vert v_i^M)}$	...	$g(v_i^M,v^K_{-i})-x_i\frac{\tau_i^\star(v^K_{-i}\vert v_i^M)}{\tau_i^\star(v^1_{-i}\vert v_i^M)}$

First, the choice of x_i guarantees that $h_i(v) \geq 0$ for all $v \in (V_i \cup \hat{V}_i) \times V_{-i}$. It is also immediate that h_i and f are ε-close to each other, because $h_i(v_i^m+\varepsilon, v^k_{-i}) = g(v_i^m, v^k_{-i}) - h_i(v_i^m, v^k_{-i})$, and from this equation, we can also easily see that $\sum_v h_i(v) = 1$.

Next, we show that under this construction of bidders' beliefs, $\bar{R}^{\star}_{\varepsilon} \leq R^D$. This, combines with the fact that $\bar{R}^{\star}_{\varepsilon} \geq R^D$ completes the proof. Note that under $\hat{V}$, the shifted valuations $\{v_i + \varepsilon\}_{v_i \in V}$ are never realized. Therefore, the seller's problem under the class of dominant-strategy mechanisms are defined entirely on V, as follows:

$$
\begin{aligned}
(P^D) \qquad \bar{R}^D = \sup_{(q,p)} \quad & \mathbb{E}_{v\sim g}[\sum_i p_i(v)] \equiv \sum_{v\in V}\sum_i p_i(v)g(v) \\
s.t. \quad & \forall i, \quad \forall v_i, v_i' \in V_i, \quad \forall v_{-i} \in V_{-i}: \\
& v_i q_i(v) - p_i(v) \geq 0 \\
& v_i q_i(v) - p_i(v) \geq v_i q_i(v_i', v_{-i}) - p_i(v_i', v_{-i}) \\
& q_i(v) \geq 0; \quad \sum_i q_i(v) \leq 1
\end{aligned}
$$

Let $\{q^D(v), p^D(v)\}_v$ denote the solution of (P^D). By a standard result, only local downward IC constraints and IR constraints for the lowest type bind. Therefore, there exist multipliers $\{\lambda_i^D(v)\}_v$ associated with those constraints such that $\{q^D(v), p^D(v), \lambda_i^D(v)\}_v$ maximizes the following Lagrangian function:

$$
\begin{aligned}
\mathcal{L}^D \equiv & \sum_{i,v} p_i(v)g(v) + \sum_{i,v_{-i}} \lambda_i^D(v_i^1, v_{-i})\left[v_i^1 q_i(v_i^1, v_{-i}) - p_i(v_i^1, v_{-i})\right] \\
& + \sum_{i,v_{-i}} \sum_{v_i^m \geq v_i^2} \lambda_i^D(v_i^m, v_{-i}) \Big[\left[v_i^m q_i(v_i^m, v_{-i}) - p_i(v_i^m, v_{-i})\right] - \left[v_i^m q_i(v_i^{m-1}, v_{-i}) - p_i(v_i^{m-1}, v_{-i})\right]\Big]
\end{aligned}
$$

over the domain $(q,p) \in \mathbb{Q} \times \mathbb{R}$ where $\mathbb{Q} \equiv \{q_i(v) \geq 0; \sum_i q_i(v) \leq 1\}$.

Note that there are no restrictions imposed on payments. Therefore, at optimum:

$$
\begin{aligned}
& \frac{\partial \mathcal{L}^D}{\partial p(v_i^M, v_{-i})} = 0 \Leftrightarrow \lambda_i^D(v_i^M, v_{-i}) = g(v_i^M, v_{-i}), \\
& \frac{\partial \mathcal{L}^D}{\partial p(v_i^m, v_{-i})} = 0 \Leftrightarrow \lambda_i^D(v_i^m, v_{-i}) = \lambda_i^D(v_i^{m+1}, v_{-i}) + g(v_i^m, v_{-i}) \quad \forall 1 \leq m < M
\end{aligned}
$$

Thus, we have for all (v_i^m, v_{-i}):

$$\lambda_i^D(v_i^m, v_{-i}) = \sum_{\hat{v}_i \geq v_i^m} g(\hat{v}_i, v_{-i}) \tag{1}$$

Similarly, the seller's problem under the class of Bayesian-strategy mechanisms is also defined entirely on V, as follows:

$$\begin{aligned}
(P^B) \quad \bar{R}_\varepsilon^\star = \sup_{(q,p)} \quad & \mathbb{E}_{v\sim g}[\sum_i p_i(v)] \equiv \sum_{v\in V}\sum_i p_i(v)g(v) \\
s.t. \quad & \forall i, \quad \forall v_i, v_i' \in V_i, \quad \forall v_{-i} \in V_{-i}: \\
& \sum_{v_{-i}} h_i(v_{-i}|v_i)[v_i q_i(v) - p_i(v)] \geq 0 \\
& \sum_{v_{-i}} h_i(v_{-i}|v_i)[v_i q_i(v) - p_i(v)] \geq \sum_{v_{-i}} h_i(v_{-i}|v_i)[v_i q_i(v_i', v_{-i}) - p_i(v_i', v_{-i})] \\
& q_i(v) \geq 0; \quad \sum_i q_i(v) \leq 1
\end{aligned}$$

Consider its relaxed problem, denoted by (RP^B) where only local downward IC constraints and IR constraints for the lowest type are considered and hence, must be binding.

Moreover, recall our construction for i's prior:

$$h_i(v_{-i}|v_i^m) = \frac{h_i(v_{-i}, v_i^m)}{\sum_{v_{-i}} h_i(v_{-i}, v_i^m)} \equiv \frac{x_i \dfrac{\tau_i^\star(v_{-i}^k|v_i^m)}{\tau_i^\star(v_{-i}^1|v_i^m)}}{\sum_{v_{-i}} x_i \dfrac{\tau_i^\star(v_{-i}^k|v_i^m)}{\tau_i^\star(v_{-i}^1|v_i^m)}} = \tau_i^\star(v_{-i}^k|v_i^m) = \frac{\sum_{\hat{v}_i\geq v_i^m} g(\hat{v}_i, v_{-i})}{\sum_{v_{-i}}\sum_{\hat{v}_i\geq v_i^m} g(\hat{v}_i, v_{-i})} \tag{2}$$

(1) and (2) imply that:

$$h_i(v_{-i}|v_i^m) = \frac{\lambda_i^D(v_i^m, v_{-i})}{\sum_{v_{-i}} \lambda_i^D(v_i^m, v_{-i})}$$

Therefore, each (binding) constraint in (RP^B) is a weighted sum of (binding) constraints in (P^D), with the weight being the corresponding optimal Lagrangian multiplier for the latter.[9] Then, it can be verified that the these two problems have the same values. Note that the value of (RP^B) is obviously an upper bound of that under the original problem (P^B). Hence, we obtain: $\bar{R}^{\star}_{\varepsilon} \leq R^D$. Therefore, $\bar{R}^{\star}_{\varepsilon} = R^D$.

C Global deviation in optimal dominant-strategy mechanisms with a misspecified support.

Recall the standard properties of the optimal dominant-strategy mechanism under the assumption that $g = f$:

1. All the local downward IC constraints bind, i.e, for any $k \geq 2$, any $\hat{v}_{-i}$, and any $s^k \in supp(S)$:

$$s^k q(s^k, \hat{v}_{-i}) - p(s^k, \hat{v}_{-i}) = s^k q(s^{k-1}, \hat{v}_{-i}) - p(s^{k-1}, \hat{v}_{-i})$$

   where $s^{k-1} = \max\{s \in supp(S) \mid s < s^k\}$.

2. Allocation is monotone, i.e., $q_i(s^k, v_{-i}) \leq q_i(s^{k'}, v_{-i})$ if $k < k'$.

If it is possible that f and g are (ε-close to but) different from each other, then global deviations would typically be relevant.

Proposition 1. *Fix ε and g. In the optimal dominant-strategy mechanism assuming g is the true prior, there exists f that is ε-close to g such that, if $v \sim f$, then a bidder does not find it optimal to report the value that is closest to his true valuation.*

[9] Hien Pham thanks Benjamin Brooks for his explanation on this point.

Proof. Let $S = \text{supp}(g)$. Let f be such that some i's value $v_i = s^k - \varepsilon$ is supported. Then, he prefers reporting s^{k-1} to reporting s^k, even though s^k is closer to v_i than s^{k-1}. This is because:

$$\begin{aligned}(s^k - \varepsilon)q(s^k, \hat{v}_{-i}) - p(s^k, \hat{v}_{-i}) &= s^k q(s^k, \hat{v}_{-i}) - p(s^k, \hat{v}_{-i}) - \varepsilon q(s^k, \hat{v}_{-i}) \\ &\leq s^k q(s^{k-1}, \hat{v}_{-i}) - p(s^{k-1}, \hat{v}_{-i}) - \varepsilon q(s^{k-1}, \hat{v}_{-i}) \\ &= (s^k - \varepsilon)q(s^{k-1}, \hat{v}_{-i}) - p(s^{k-1}, \hat{v}_{-i})\end{aligned}$$

where the inequality follows from the local $\text{IC}_{k,k-1}$ constraint and the monotonicity constraint. Moreover, if $q(s^k, \hat{v}_{-i}) > q(s^{k-1}, \hat{v}_{-i})$, the inequality then becomes strict. That is:

$$(s^k - \varepsilon)q(s^k, \hat{v}_{-i}) - p(s^k, \hat{v}_{-i}) < (s^k - \varepsilon)q(s^{k-1}, \hat{v}_{-i}) - p(s^{k-1}, \hat{v}_{-i})$$

Consequently, the agent whose value is $v_i = s^k - \varepsilon$ strictly prefer to report his valuation as s^{k-1} instead of his closest type s^k. □

D Proof of Theorem 2

Let $(q^\star(\cdot), p^\star(\cdot))$ represent the optimal dominant-strategy mechanism under the assumption that g is the true prior. Let $V = \text{supp}(g)$, and let $V_i = \{v_i \in \mathbb{R} \mid \exists v_{-i};\ (v_i, v_{-i}) \in V\}$ denote its i-th coordinate. We also denote $v_i^+ \equiv \min\{s \in V_i \mid s > v_i\}$ and $v_i^- \equiv \max\{s \in V_i \mid s < v_i\}$.

Fix $\delta > 0$, which is sufficiently small. The *δ-transfer reduction mechanism* of $(q^\star(\cdot), p^\star(\cdot))$ has the same message space and the winning-probability function as the optimal dominant-strategy mechanism, but the price is smaller by δ.

For each $v_i \in V_i$, truth-telling is still dominant-strategy incentive compatible, but now in a stronger sense: for bidder i whose value is δ-close to $v_i \in V_i$, it is dominant for him to report v_i in the δ-transfer-reduction mechanism. Note that under the original mechanism, for all $\sigma(v_{-i})$ and $v_i' < v_i$, we have $q_i^\star(v_i, \sigma(v_{-i})) \geq q_i^\star(v_i', \sigma(v_{-i}))$ and:

$$v_i q_i^\star(v_i, \sigma(v_{-i})) - p_i^\star(v_i, \sigma(v_{-i})) \geq v_i q_i^\star(v_i', \sigma(v_{-i})) - p_i^\star(v_i', \sigma(v_{-i}))$$

which means:

$$\begin{aligned} &(v_i - \delta) q_i^\star(v_i, \sigma(v_{-i})) - [p_i^\star(v_i, \sigma(v_{-i})) - \delta q_i^\star(v_i, \sigma(v_{-i}))] \\ \geq &(v_i - \delta) q_i^\star(v_i', \sigma(v_{-i})) - [p_i^\star(v_i', \sigma(v_{-i})) - \delta q_i^\star(v_i', \sigma(v_{-i}))] \end{aligned}$$

By single crossing property and $\hat{v}_i \geq v_i - \delta$ ($\hat{v}_i$ is δ-close to $v_i \in V_i$), we thus have:

$$\hat{v}_i q_i^\star(v_i, \sigma(v_{-i})) - [p_i^\star(v_i, \sigma(v_{-i})) - \delta q_i^\star(v_i, \sigma(v_{-i}))] \geq \hat{v}_i q_i^\star(v_i', \sigma(v_{-i})) - [p_i^\star(v_i', \sigma(v_{-i})) - \delta q_i^\star(v_i', \sigma(v_{-i}))]$$

for all $\sigma(v_{-i})$ and $v_i' < v_i$, i.e.,

$$v_i = \arg\max_{v_i' \leq v_i} \Big[\hat{v}_i q_i^\star(v_i', \sigma(v_{-i})) - [p_i^\star(v_i', \sigma(v_{-i})) - \delta q_i^\star(v_i', \sigma(v_{-i}))] \Big] \quad \forall \sigma(v_{-i}) \tag{3}$$

Note also that under the original mechanism, for all $\sigma(v_{-i})$ and $v_i' > v_i^+$, we have $q_i^\star(v_i', \sigma(v_{-i})) \geq q_i^\star(v_i^+, \sigma(v_{-i})) \geq q_i^\star(v_i, \sigma(v_{-i}))$, and:

$$v_i^+ q_i^\star(v_i^+, \sigma(v_{-i})) - p_i^\star(v_i^+, \sigma(v_{-i})) = v_i^+ q_i^\star(v_i, \sigma(v_{-i})) - p_i^\star(v_i, \sigma(v_{-i}))$$

which means:

$$\begin{aligned} &(v_i^+ - \delta) q_i^\star(v_i^+, \sigma(v_{-i})) - [p_i^\star(v_i^+, \sigma(v_{-i})) - \delta q_i^\star(v_i^+, \sigma(v_{-i}))] \\ = &(v_i^+ - \delta) q_i^\star(v_i, \sigma(v_{-i})) - [p_i^\star(v_i, \sigma(v_{-i})) - \delta q_i^\star(v_i, \sigma(v_{-i}))] \end{aligned}$$

By single crossing property and $v_i^+ - \delta \geq \hat{v}_i$ ($\hat{v}_i$ is δ-close to $v_i \in V_i$), this implies that for all $\sigma(v_{-i})$:

$$\begin{aligned} &\hat{v}_i q_i^\star(v_i^+, \sigma(v_{-i})) - [p_i^\star(v_i^+, \sigma(v_{-i})) - \delta q_i^\star(v_i^+, \sigma(v_{-i}))] \\ \leq &\hat{v}_i q_i^\star(v_i, \sigma(v_{-i})) - [p_i^\star(v_i, \sigma(v_{-i})) - \delta q_i^\star(v_i, \sigma(v_{-i}))] \end{aligned} \tag{4}$$

Moreover, for all $\sigma(v_{-i})$ and $v_i' > v_i^+$:

$$v_i^+ q_i^\star(v_i', \sigma(v_{-i})) - p_i^\star(v_i', \sigma(v_{-i})) \leq v_i^+ q_i^\star(v_i^+, \sigma(v_{-i})) - p_i^\star(v_i^+, \sigma(v_{-i}))$$

which means:

$$\begin{aligned} &(v_i^+ - \delta) q_i^\star(v_i', \sigma(v_{-i})) - [p_i^\star(v_i', \sigma(v_{-i})) - \delta q_i^\star(v_i', \sigma(v_{-i}))] \\ \leq &(v_i^+ - \delta) q_i^\star(v_i^+, \sigma(v_{-i})) - [p_i^\star(v_i^+, \sigma(v_{-i})) - \delta q_i^\star(v_i^+, \sigma(v_{-i}))] \end{aligned}$$

By single crossing property and $v_i^+ - \delta \geq \hat{v}_i$ ($\hat{v}_i$ is δ-close to $v_i \in V_i$), we thus have for all $\sigma(v_{-i})$ and $v_i' > v_i^+$:

$$\begin{aligned} &\hat{v}_i q_i^\star(v_i', \sigma(v_{-i})) - [p_i^\star(v_i', \sigma(v_{-i})) - \delta q_i^\star(v_i', \sigma(v_{-i}))] \\ \leq &\hat{v}_i q_i^\star(v_i^+, \sigma(v_{-i})) - [p_i^\star(v_i^+, \sigma(v_{-i})) - \delta q_i^\star(v_i^+, \sigma(v_{-i}))] \end{aligned} \tag{5}$$

Combining (4) and (5), we obtain:

$$v_i = \arg\max_{v_i' \geq v_i} \Big[\hat{v}_i q_i^\star(v_i', \sigma(v_{-i})) - [p_i^\star(v_i', \sigma(v_{-i})) - \delta q_i^\star(v_i', \sigma(v_{-i}))] \Big] \quad \forall \sigma(v_{-i}) \tag{6}$$

Then, (3) and (6) imply that for bidder i whose value is δ-close to $v_i \in V_i$, it is dominant for him to report v_i in the δ-transfer-reduction mechanism. We take $\delta = \varepsilon$ then. Although we omit the details, it can also be shown that his ex post individual rationality is satisfied.

By construction, the ε-transfer-reduction mechanism collects the same amount of transfer from each type of each agent less at most ε. Therefore, *if* $g = f$, then the expected revenue in the ε-transfer-reduction mechanism, denoted by $R'_\varepsilon(g)$, is not lower than $R^D - N\varepsilon$:

$$R'_\varepsilon(g) \geq R^D - N\varepsilon.$$

Even if f is different from g, it remains true that each bidder with each type finds it dominant to report his closest type in the same ε-transfer-reduction mechanism. Therefore, denoting by $R'_\varepsilon(f)$ the expected revenue of the same ε-transfer-reduction mechanism but with distribution f, by continuity we obtain:

$$\lim_{\varepsilon\to 0} |R'_\varepsilon(g) - R'_\varepsilon(f)| = 0,$$

and therefore:

$$\lim_{\varepsilon\to 0} |R'_\varepsilon(g) - \inf_{f|\varepsilon\text{-close to } g} R'_\varepsilon(f)| = 0,$$

By Theorem 1:

$$R^D = R^\star_\varepsilon \geq \inf_{f|\varepsilon\text{-close to } g} R'_\varepsilon(f)$$

Therefore:

$$0 \leq R^D - \inf_{f|\varepsilon\text{-close to } g} R'_\varepsilon(f) \leq R'_\varepsilon(g) + N\varepsilon - \inf_{f|\varepsilon\text{-close to } g} R'_\varepsilon(f),$$

where the right-hand side converges to 0 as $\varepsilon \to 0$, implying:

$$\inf_{f|\varepsilon\text{-close to } g} R'_\varepsilon(f) \to R^D,$$

as $\varepsilon \to 0$. We complete the proof by noticing that $R^D_\varepsilon \in [\inf_{f|\varepsilon\text{-close to } g} R'_\varepsilon(f), R^D]$.

References

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

Bergemann, D., and K. Schlag (2011): "Robust Monopoly Pricing," *Journal of Economic Theory*, 146(6), 2527–2543.

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, 139–166.

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

Chung, K., and J. Ely (2007): "Foundations of Dominant-Strategy Mechanisms," *The Review of Economic Studies*, 74(2), 447–476.

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

Eliaz, K., and R. Spiegler (2008): "Consumer optimism and price discrimination," *Theoretical Economics*, 3(4), 459–497.

Grubb, M. D. (2009): "Selling to overconfident consumers," *American Economic Review*, 99(5), 1770–1807.

Jehiel, P., M. Meyer-ter Vehn, and B. Moldovanu (2012): "Locally Robust Implementation and Its Limits," *Journal of Economic Theory*, 147(6), 2439–2452.

Jehiel, P., M. M. ter Vehn, B. Moldovanu, and W. R. Zame (2006): "The Limits of Ex Post Implementation," *Econometrica*, 74(3), 585–610.

Lopomo, G., L. Rigotti, and C. Shannon (2021): "Uncertainty in Mechanism Design," *working paper*.

——— (2022): "Uncertainty and robustness of surplus extraction," *Journal of Economic Theory*, 199, 105088.

Madarász, K., and A. Prat (2017): "Sellers with Misspecified Models," *The Review of Economic Studies*, 84(2), 790–815.

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

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