Theoretical Economics 11 (2016), 971–1004 1555-7561/20160971

Mechanism design with maxmin agents: Theory and an application to bilateral trade

Alexander Wolitzky
Department of Economics, MIT

This paper studies mechanism design when agents are maxmin expected utility maximizers. A first result gives a general necessary condition for a social choice rule to be implementable. The condition combines an inequality version of the standard envelope characterization of payoffs in quasilinear environments with an approach for relating agents' maxmin expected utilities to their objective expected utilities under any common prior. The condition is then applied to give an exact characterization of when efficient trade is possible in the bilateral trading problem of Myerson and Satterthwaite (1983), under the assumption that agents know little beyond each other's expected valuation of the good (which is the information structure that emerges when agents are uncertain about each other's ability to acquire information). Whenever efficient trade is possible, it may be implemented by a relatively simple double auction format. Sometimes, an extremely simple reference price rule can also implement efficient trade.

Keywords. Mechanism design, maxmin, ambiguity aversion, bilateral trade, Myerson–Satterthwaite.

JEL classification. D81, D82.

1. Introduction

"Robustness" has been a central concern in game theory and mechanism design since at least the celebrated argument of Wilson (1989). The Wilson doctrine is usually interpreted as calling for mechanisms that perform well in a wide range of environments. However, there is also growing and complementary interest in robustness concerns on the part of economic agents instead of (or in addition to) on the part of the mechanism designer; that is, in asking what mechanisms are desirable when agents use "robustly optimal" strategies. This paper pursues this question in the case where agents are maxmin expected utility (MMEU) maximizers (Gilboa and Schmeidler 1989), which

Alexander Wolitzky: wolitzky@mit.edu
For helpful comments, I thank Pierpaolo Battigalli, Aaron Bodoh-Creed, Subir Bose, Gabe Carroll, Glenn Ellison, Aytek Erdil, Ben Golub, Faruk Gul, Fuhito Kojima, Massimo Marinacci, Paul Milgrom, Stephen Morris, Suresh Mutuswami, Phil Reny, Ilya Segal, Andy Skrzypacz, Joel Sobel, Juuso Toikka, several anonymous referees, and seminar participants at Brown, Collegio Carlo Alberto, Columbia, Indiana, Leicester, Northwestern, NYU, Princeton, Simon Fraser, Stanford, Toronto, UBC, USC, UT Austin, Washington University, and the 2014 Cowles Foundation Conference on Economic Theory. I thank Jason Huang for excellent research assistance.

DOI: 10.3982/TE2089

is perhaps the best-established model of robust decision-making under uncertainty, as well as the model most commonly adopted in prior studies of mechanism design with robustness concerns on the part of agents. In particular, the paper develops a general necessary condition for a social choice rule to be implementable, and applies this condition to give an exact characterization of when efficient trade is possible in the classical bilateral trade setting of Myerson and Satterthwaite (1983).

The necessary condition for implementation generalizes a well known necessary condition in the Bayesian independent private values setting, namely that the expected social surplus must exceed the expected sum of information rents left to the agents, as given by an envelope theorem. That this condition has any analogue with maxmin agents is rather surprising, for two reasons. First, the usual envelope characterization of payoffs need not hold with maxmin agents. Second, and more importantly, a maxmin agent's subjective belief about the distribution of opposing types depends on her own type. This is also the situation with Bayesian agents and *correlated* types, where results are quite different from those in the classical independent types case (Crémer and McLean 1985, 1988, McAfee and Reny 1992).

The derivation of the necessary condition (Theorem 1) addresses both of these issues. For the first, I rely on an inequality version of the standard envelope condition that does hold with maxmin agents. For the second, I note that, by definition, an agent's maxmin expected utility is lower than her expected utility under any belief she finds possible. This implies that the sum of agents' maxmin expected utilities is lower than the sum of their "objective" expected utilities under any possible common prior, which in turn equals the expected social surplus under that prior (for a budget-balanced mechanism). Hence, a necessary condition for a social choice rule to be implementable is that the resulting expected social surplus exceeds the expected sum of information rents for *any* possible common prior; that is, for any prior with marginals that the agents find possible.

The second part of the paper applies this necessary condition to give an exact characterization of when efficient bilateral trade is implementable, under the assumption that the agents know each other's expected valuation of the good (as well as bounds on the valuations), but little else. As explained below, this is the information structure that emerges when agents have a (unique) common prior on values at an ex ante stage and are maxmin about how the other agent might acquire information before participating in the mechanism. In this setting, the assumption of maxmin behavior may be an appealing alternative to the Bayesian approach of specifying a prior over the set of experiments that the other agent may have access to, especially when this set is large (e.g., consists of all possible experiments) or the agents' interaction is one shot. Furthermore, the great elegance of Myerson and Satterthwaite's theorem and proof suggests that their setting may be one where relaxing the assumption of a unique common prior is particularly appealing.[1]

[1]This is in line with Gilboa's exhortation in his monograph on decision-making under uncertainty to "[consider] the MMEU model when a Bayesian result seems to crucially depend on the existence of a unique, additive prior, which is common to all agents. When you see that, in the course of some proof, things cancel out too neatly, this is the time to wonder whether introducing a little bit of uncertainty may provide more realistic results" (Gilboa 2009, p. 169).

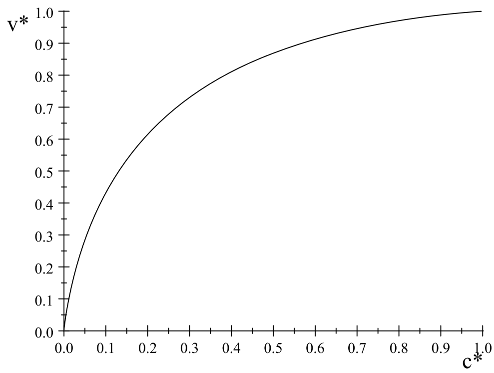

FIGURE 1. In the bilateral trade setting, efficient trade is possible in the region below the curve and impossible in the region above it.

The second main result (Theorem 2) shows that the Myerson–Satterthwaite theorem sometimes continues to hold when agents are maxmin about each other's information acquisition technology—but sometimes not. In the simplest bilateral trade setting, where the range of possible seller costs and buyer values is $[0, 1]$, the average seller cost is c^*, and the average buyer value is v^*, Figure 1 indicates the combination of parameters (c^*, v^*) for which an efficient, maxmin incentive compatible, interim individually rational, and weakly budget balanced mechanism exists. Above the curve, the formula for which is

$$\frac{c^*}{1-c^*}\log\left(1+\frac{1-c^*}{c^*}\right)+\frac{1-v^*}{v^*}\log\left(1+\frac{v^*}{1-v^*}\right)=1,$$

the Myerson–Satterthwaite theorem persists, despite the lack of a unique common prior or independent types. Below the curve, the Myerson–Satterthwaite theorem fails.

I call the mechanism that implements efficient trade for all parameters below the curve in Figure 1 the $\alpha_i(\theta_i)$ *double auction*. It is so-called because when a type θ_i agent and a type θ_j agent trade, the type θ_i agent receives a share $\alpha_i(\theta_i)$ of the gains from trade that depends only on her own type and not on her opponent's. The $\alpha_i(\theta_i)$ double auction has the property that an agent's worst-case belief is the belief that minimizes the probability that strict gains from trade exist; this may be seen to be the belief that her opponent's type always takes on either the most favorable value for which there are no gains from trade or the most favorable value possible. If an agent misreports her type to try to get a better price, the requirement that her opponent's average value is fixed forces the deviator's worst-case belief to put more weight on the less favorable of these values, which reduces her expected probability of trade. The share $\alpha_i(\theta_i)$ is set so that this first-order cost in terms of the probability of trade exactly offsets the first-order benefit in terms of price, which makes the $\alpha_i(\theta_i)$ double auction incentive compatible for maxmin agents.[2] Finally, the $\alpha_i(\theta_i)$ double auction is weakly budget balanced if and

[2]In contrast, the cost of shading one's report in terms of foregone gains from trade would be second order for a Bayesian, as both the probability that shading results in a missed opportunity to trade and the foregone gains from trade conditional on missing a trading opportunity would be small.

only if $\alpha_1(\theta_1) + \alpha_2(\theta_2) \leq 1$ for all θ_1, θ_2; that is, if and only if the shares that must be left to the two agents sum to less than 1. This inequality holds in precisely the region below the curve in Figure 1.

I also derive some additional results in the bilateral trade setting. Most notably, I show that if the average types of the two agents do not have gains from trade with each other (e.g., if the pair (c^*, v^*) lies below the 45° line in Figure 1), then efficient trade can be implemented with an extremely simple mechanism, which I call a *reference rule*. A reference rule works by setting a "reference price" p^* and specifying that trade occurs at price p^* if this is acceptable to both agents, and otherwise that trade occurs (when efficient) at the reservation price of the agent who refuses to trade at p^*. This result thus illustrates a case where introducing robustness concerns on the part of agents leads simple mechanisms to satisfy desirable mechanism design criteria.

This paper joins a growing literature on games and mechanisms with maxmin agents or with agents who follow "robust" decision rules more generally. In contrast to much of this literature, the current paper shares the following important features of classical Bayesian mechanism design: (i) the implementation concept is (partial) Nash implementation; (ii) the only source of uncertainty in the model concerns exogenous random variables, namely other agents' types; (iii) for Theorem 1, the model admits the possibility of a unique common prior as a special case. Several recent papers derive permissive implementability results with maxmin agents by relaxing these assumptions, in contrast to the relatively restrictive necessary condition of Theorem 1.

Bose and Daripa (2009), Bose and Mutuswami (2012), and Bose and Renou (2014) relax (i) by considering dynamic mechanisms that exploit the fact that maxmin agents may be time-inconsistent. A central feature of their approach is that agents cannot commit to strategies, so they do not obtain implementation in Nash equilibrium. Their approach also relies on taking a particular position on how maxmin agents update their beliefs, an issue that does not arise here. Di Tillio et al. (2014) and Bose and Renou (2014) relax (ii) by assuming that agents are maxmin over uncertain aspects of the mechanism itself. This lets the designer extract the agents' information by introducing "bait" provisions into the mechanism. The mechanisms considered in these four papers are undoubtedly interesting and may be appealing in particular applications. However, they arguably rely on a more thoroughgoing commitment to maxmin behavior than does the current paper (agents must be time-inconsistent or must be maxmin over endogenous random variables). Even if one accepts this commitment, it still seems natural to ask what is possible in the more "standard" case where (i) and (ii) are satisfied.

De Castro and Yannelis (2010) relax (iii) by assuming that agents' beliefs are completely unrestricted, and they find that efficient social choice rules are then always implementable. This is consistent with Theorem 1, as with completely unrestricted beliefs agents can always expect the worst possible allocation, which implies that the necessary condition of Theorem 1 is vacuously satisfied. For example, in the bilateral trade setting, efficient trade is always implementable, as agents are always certain that they will not trade and are therefore willing to reveal their types. Thus, De Castro and Yannelis show that ambiguity aversion can soften the Myerson–Satterthwaite impossible result—consistent with Theorem 2—but they do so only under the rather extreme assumption of completely unrestricted beliefs.

Finally, Bose et al. (2006) and Bodoh-Creed (2012) satisfy (i), (ii), and (iii).[3] Their results are discussed below, but the main differences are that neither of these papers derives a general necessary condition for implementability like Theorem 1, and their treatment of applications focuses not on efficiency, but on revenue maximization. Importantly, this revenue maximization is conducted with respect to the mechanism designer's "true" prior, whereas in my model there is no notion of a true prior and the designer is simply a stand-in for all the various games the agents could play among themselves. For example, Bodoh-Creed does consider an application to bilateral trade, but he investigates the minimum expected budget deficit required to implement efficient trade (from the designer's perspective), rather than whether efficient trade is possible with ex post budget balance.[4]

The paper proceeds as follows. Section 2 presents the model. Section 3 gives the general necessary condition for implementation. Section 4 applies this condition to characterize when efficient bilateral trade is implementable. Section 5 contains additional results in the bilateral trade setting, including the results on implementation with reference rules. Section 6 concludes. The Appendix contains omitted proofs and auxiliary results.

2. Model

Agents and Preferences: A group N of n agents must make a social choice from a bounded set of alternatives $Y \subseteq \mathbb{R}^n$. Each agent i has a one-dimensional type $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i] = \Theta_i \subseteq \mathbb{R}$. Agents have quasilinear utility. In particular, if alternative $y = (y_1, \dots, y_n)$ is selected and a type θ_i agent receives transfer t_i, her payoff is $\theta_i y_i + t_i$.[5]

Agent i's type is her private information. In addition, each agent i has a *set of possible beliefs* Φ_{-i} about her opponents' types, where Φ_{-i} is an arbitrary nonempty subset of $\Delta(\Theta_{-i})$, the set of Borel measures ϕ_{-i} on Θ_{-i}. (Throughout, probability measures are denoted by ϕ, and the corresponding cumulative distribution functions are denoted by F.) Each agent i evaluates her expected utility with respect to the worst possible distribution of her opponents' types among those distributions in Φ_{-i}; that is, the agents are maxmin optimizers.

[3] Lopomo et al. (2014) also satisfy (i), (ii), and (iii), but consider agents with incomplete preferences as in Bewley (2002) rather than maxmin preferences. There are two natural versions of incentive compatibility in their model, which bracket maxmin incentive compatibility (and Bayesian incentive compatibility) in terms of strength. They show that the stronger of their notions of incentive compatibility is often equivalent to ex post incentive compatibility (whereas maxmin incentive compatibility is not), and that full extraction of information rents is generically possible under the weaker of their notions and is sometimes possible under the stronger one.

[4] The literature on mechanism design with *risk-averse* agents is more tangentially related to the current paper. Chatterjee and Samuelson (1983) and Garratt and Pycia (2015) propose mechanisms for efficient bilateral trade with risk-averse agents. In contrast, I maintain the assumption that utility is quasilinear. The mechanisms I propose bear little resemblance to those proposed for risk-averse agents.

[5] The assumption that utility is multiplicative in θ_i and y_i is for simplicity. One could instead assume that utility equals $v_i(y, \theta_i) + t_i$ for some absolutely continuous and equidifferentiable family of functions $\{v_i(y, \cdot)\}$, as in Milgrom and Segal (2002) or Bodoh-Creed (2012).

Mechanisms: A *direct mechanism* (y, t) consists of a measurable allocation rule $y : \Theta \to Y$ and a measurable and bounded transfer rule $t : \Theta \to \mathbb{R}^n$. Given a mechanism (y, t), let

$$U_i(\hat{\theta}_i, \theta_{-i}; \theta_i) = \theta_i y_i(\hat{\theta}_i, \theta_{-i}) + t_i(\hat{\theta}_i, \theta_{-i})$$
$$U_i(\hat{\theta}_i, \phi_{-i}; \theta_i) = E^{\phi_{-i}}[U_i(\hat{\theta}_i, \theta_{-i}; \theta_i)]$$
$$U_i(\theta_i) = \inf_{\phi_{-i} \in \Phi_{-i}} U_i(\theta_i, \phi_{-i}; \theta_i).$$

Thus, $U_i(\hat{\theta}_i, \theta_{-i}; \theta_i)$ is agent i's utility from reporting type $\hat{\theta}_i$ against opposing type profile θ_{-i} given true type θ_i, $U_i(\hat{\theta}_i, \phi_{-i}; \theta_i)$ is agent i's expected utility from reporting type $\hat{\theta}_i$ against belief ϕ_{-i} given true type θ_i, and $U_i(\theta_i)$ is agent i's worst-case expected utility from reporting her true type θ_i.[6]

A distinguishing feature of this paper is the notion of incentive compatibility employed, which I call *maxmin incentive compatibility*. A mechanism is maxmin incentive compatible (MMIC) if

$$\theta_i \in \arg\max_{\hat{\theta}_i \in \Theta_i} \inf_{\phi_{-i} \in \Phi_{-i}} U_i(\hat{\theta}_i, \phi_{-i}; \theta_i) \quad \text{for all } \theta_i \in \Theta_i, i \in N. \tag{1}$$

I restrict attention to MMIC direct mechanisms throughout the paper. This is without loss of generality under the assumption that agents cannot hedge against ambiguity by randomizing, in that an agent's utility from playing a mixed strategy $\mu_i \in \Delta(\Theta_i)$ is $E^{\mu_i} \inf_{\phi_{-i} \in \Phi_{-i}} U_i(\hat{\theta}_i, \phi_{-i}; \theta_i)$ rather than $\inf_{\phi_{-i} \in \Phi_i} E^{\mu_i} U_i(\hat{\theta}_i, \phi_{-i}; \theta_i)$. Under this "no-hedging" assumption, the proof of the revelation principle is completely standard.[7]

A brief aside on the no-hedging assumption: While the alternative is also reasonable, the no-hedging assumption is the standard one in decision theory. In particular, the uncertainty aversion axiom of Schmeidler (1989) and Gilboa and Schmeidler (1989) says that the agent likes mixing ex post (i.e., state by state); mixing over acts ex ante does not affect her utility. In addition, as noted by Raiffa (1961), if agents could hedge with randomization, then one would not observe the Ellsberg paradox or other well documented, ambiguity-averse behavior. The no-hedging assumption is also standard in the literature on mechanism design with maxmin agents (e.g., Bose et al. 2006, De Castro and Yannelis 2010, Bodoh-Creed 2012, Di Tillio et al. 2014).[8]

In addition to MMIC, I consider the following standard mechanism design criteria.

[6]The term "worst case" is only used heuristically in this paper, but the meaning is generally that if $\min_{\phi_{-i} \in \Phi_{-i}} U_i(\hat{\theta}_i, \phi_{-i}; \theta_i)$ exists, then a minimizer is a worst-case belief; while if the minimum does not exist (which is possible, as $U_i(\hat{\theta}_i, \phi_{-i}; \theta_i)$ may not be continuous in ϕ_{-i} and Φ_{-i} may not be compact), then a limit point of a sequence $\{\phi_{-i}\}$ that attains the infimum is a worst-case belief.

[7]Under the solution concept of Nash equilibrium. In particular, there is no strategic uncertainty or "higher order ambiguity" (as in Ahn 2007). See the working paper version of this paper (Wolitzky 2016) for further details.

[8]However, Agranov and Ortoleva (forthcoming) present experimental evidence that sometimes agents do display a strict preference for randomization. Models of such preferences include Machina (1985), Cerreia-Vioglio et al. (2015), and Fudenberg et al. (2015). Saito (2015) axiomatizes a utility function that identifies an agent's belief that randomization hedges ambiguity.

- *Ex Post Efficiency (EF)*: For all $\theta \in \Theta$ $y(\theta) \in \arg\max_{y \in Y} \sum_i \theta_i y_i$.
- *Interim Individual Rationality (IR)*: For all $\theta_i \in \Theta_i$ $U_i(\theta_i) \geq 0$.
- *Ex Post Weak Budget Balance (WBB)*: For all $\theta \in \Theta$ $\sum_i t_i(\theta) \leq 0$.
- *Ex Post Strong Budget Balance (SBB)*: For all $\theta \in \Theta$ $\sum_i t_i(\theta) = 0$.

Efficiency is self-explanatory. Interim individual rationality is imposed with respect to agents' own worst-case beliefs; it also happens that all results in the paper continue to hold with ex post individual rationality (i.e., $U_i(\theta_i, \theta_{-i}; \theta_i) \geq 0$ for all $\theta_i \in \Theta_i$, $\theta_{-i} \in \Theta_{-i}$). The ex post version of budget balance seems appropriate in the absence of a "true" prior distribution; as indicated above, this focus on ex post budget balance is an important point of contrast to the otherwise closely related papers of Bose et al. (2006) and Bodoh-Creed (2012). The difference between weak and strong budget balance is that with weak budget balance, the mechanism is allowed to run a surplus. For purposes of comparison with the results of Section 4, recall that the standard Myerson–Satterthwaite theorem requires only (ex ante) weak budget balance. Also note that since MMIC is a weaker condition than dominant strategy incentive compatibility, an efficient allocation rule can always be implemented with a Vickrey–Clarke–Groves (VCG) mechanism if budget balance is not required.

An allocation rule y is *maxmin implementable* if there exists a transfer rule t such that the mechanism (y, t) satisfies MMIC, IR, and WBB.

3. Necessary conditions for implementation

I begin with a general necessary condition for maxmin implementation, which generalizes a standard necessary condition for Bayesian implementation with independent private values. In an independent private values environment with common prior distribution F, it is well known that an allocation rule y is Bayesian implementable only if the expected social surplus under y exceeds the expected information rents that must be left for the agents so as to satisfy incentive compatibility. It follows from standard arguments (e.g., Myerson 1981) that this condition may be written as

$$\sum_i \left(\int_{\theta \in \Theta} \theta_i y_i(\theta)\, d\phi \right) \geq \sum_i \left(\int_{\theta_i \in \Theta_i} (1 - F_i(\theta_i)) y_i(\theta_i, \phi_{-i})\, d\theta_i \right), \tag{2}$$

where

$$y_i(\theta_i, \phi_{-i}) = E^{\phi_{-i}}[y_i(\theta_i, \theta_{-i})]$$

(recall that ϕ is the measure corresponding to cumulative distribution function F). I will show that a similar condition is necessary for maxmin implementation, despite the lack of independent types (in that an agent's worst-case belief over her opponent's types depends on her own type) or a unique common prior. Intuitively, the required condition will be that (2) holds *for all distributions F with marginals that the agents find possi-*

ble, with the modification that, on the right-hand side of (2), the expected information rents under F are replaced by the expectation under F of type θ_i's "minimum possible" information rent.

To formalize this, given a measure $\phi \in \Delta(\Theta)$, let ϕ_S denote its marginal with respect to Θ_S for $S \subseteq N$. Let Φ^* be the set of product measures $\phi \in \prod_{i \in N} \Delta(\Theta_i)$ such that $\phi_{-i} \in \Phi_{-i}$ for all $i \in N$.

Some examples may clarify this definition.

- If $n = 2$, then $\Phi^* = \Phi_1 \times \Phi_2$ (where $\Phi_i \equiv \Phi_{-j}$).
- Suppose the set of each agent i's possible beliefs takes the form of a product $\Phi_{-i} = \prod_{j \neq i} \Phi_j^i$ for some sets of measures $\Phi_j^i \subseteq \Delta(\Theta_j)$. Then $\Phi^* = \prod_{j \in N}(\bigcap_{i \neq j} \Phi_j^i)$.
- If $n > 2$, it is possible that Φ^* is empty. For instance, take the previous example with $\bigcap_{i \neq j} \Phi_j^i = \emptyset$ for some j.

Finally, let

$$\tilde{y}_i(\theta_i) = \inf_{\phi_{-i} \in \Phi_{-i}} y_i(\theta_i, \phi_{-i}).$$

Thus, $\tilde{y}_i(\theta_i)$ is the smallest allocation that type θ_i may expect to receive.

The following result gives the desired necessary condition.

Theorem 1. *If allocation rule y is maxmin implementable, then, for every measure $\phi \in \Phi^*$,*

$$\sum_i \left(\int_{\theta \in \Theta} \theta_i y_i(\theta)\, d\phi \right) \geq \sum_i \left(\int_{\theta_i \in \Theta_i} (1 - F_i(\theta_i)) \tilde{y}_i(\theta_i)\, d\theta_i \right). \tag{3}$$

Comparing (2) and (3), (2) says that the expected social surplus under F must exceed the expected information rents, whereas (3) says that the expected social surplus must exceed the expectation of the agents' minimum possible information rents, reflecting the fact that agents' subjective expected allocations are not derived from F. In addition, (2) must hold only for the "true" distribution F (i.e., the common prior distribution), while (3) must hold for any "candidate" distribution F (i.e., any distribution in Φ^*). Furthermore, (3) is a generalization of (2), since in the case of a unique independent common prior ϕ it follows that $\Phi_{-i} = \{\phi_{-i}\}$ for all i, $\Phi^* = \{\phi\}$, and $\tilde{y}_i(\theta_i) = y_i(\theta_i, \phi_{-i})$, so (3) reduces to (2). Finally, if y is continuous, then (3) also shows that (2) changes continuously as a slight degree of ambiguity aversion is introduced into a Bayesian model.

The differences between (2) and (3) suggest that maxmin implementation is neither easier nor harder than Bayesian implementation in general and, more generally, that expanding the sets of possible beliefs Φ_{-i} can make implementation either easier or harder. In particular, expanding the sets Φ_{-i} expands Φ^*, which implies that (3) must hold for a larger set of measures ϕ. However, expanding Φ_{-i} also reduces $\tilde{y}_i(\theta_i)$ and, thus, reduces the right-hand side of (3), making (3) easier to satisfy. Indeed, Section 4 shows that efficient bilateral trade is sometimes maxmin implementable in cases where the Myerson–Satterthwaite theorem implies that it is not Bayesian implementable, which demonstrates that maxmin implementation can be easier than Bayesian implementation. But it is also easy to find examples where expanding the set of agents' possible beliefs makes implementation more difficult. For instance, take a Bayesian bilateral trade setting where all types are certain that gains from trade exist—so that efficient trade is implementable—and expand the set of possible beliefs by adding a less favorable prior for which the Myerson–Satterthwaite theorem applies. Condition (3) will then imply that efficient trade is not implementable, by exactly the same argument as in Myerson–Satterthwaite.

An important tool for proving Theorem 1 is an inequality version of the usual envelope characterization of payoffs, Lemma 1. Lemma 1 was previously derived by Bose et al. (2006, equation (8), p. 420) in the special case where Φ_i is a set of "ε contaminated" beliefs, but the proof is the same in the general case and is thus omitted.[9]

Lemma 1. *In any maxmin incentive compatible mechanism,*

$$U_i(\theta_i) \geq U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \tilde{y}_i(s)\, ds \quad \textit{for all } \theta_i \in \Theta_i. \tag{4}$$

Lemma 1 is also related to Theorem 1 of Bodoh-Creed (2012), which gives an exact characterization of payoffs using Milgrom and Segal's (2002) envelope theorem for saddle point problems. The difference comes because the maxmin problem (1) admits a saddle point in Bodoh-Creed but *not* in the present paper; one reason why is that Bodoh-Creed assumes that $y_i(\theta_i, \phi_{-i})$ is continuous in θ_i and ϕ_{-i} (his assumption A8), which may not be the case here.[10] For example, efficient allocation rules are not continuous, so Bodoh-Creed's characterization need not apply for efficient mechanisms. I discuss below how Theorem 1 may be strengthened if (1) is assumed to admit a saddle point.

With Lemma 1 in hand, the proof of Theorem 1 relates the bound on agents' subjective expected utilities in (4) to the objective social surplus on the left-hand side of (3). The key reason why this is possible is that a maxmin agent's subjective expected utility is a lower bound on her expected utility under any probability distribution she finds possible. Hence, the sum of agents' subjective expected utilities is a lower bound on the sum of their objective expected utilities under any measure $\phi \in \Phi^*$, which in turn is a lower bound on the objective expected social surplus under ϕ (if weak budget balance is satisfied). Note that this step relies crucially on the assumption that agents are maxmin optimizers; for example, it would not apply for Bayesian agents with arbitrary heterogeneous priors.

[9]Lemma 1 can also be derived as a corollary of Theorem 1 of Carbajal and Ely (2013). See also Segal and Whinston (2002) and Kos and Messner (2013) for related approaches.

[10]A careful reading of Bodoh-Creed (2012) reveals that some additional assumptions are also required for the existence of a saddle point, such as quasiconcavity assumptions. Bodoh-Creed (2014) provides an alternative derivation of his payoff characterization result under additional continuity assumptions. Neither set of assumptions is satisfied in the current setting.

Proof of Theorem 1. Suppose mechanism (y, t) satisfies MMIC, IR, and WBB. For any measure $\phi_i \in \Delta(\Theta_i)$, integrating (4) by parts yields

$$\int_{\Theta_i} U_i(\theta_i)\,d\phi_i \geq U_i(\underline{\theta}_i) + \int_{\Theta_i} (1 - F_i(\theta_i))\tilde{y}_i(\theta_i)\,d\theta_i.$$

Recall that

$$U_i(\theta_i) \leq \int_{\Theta_{-i}} (\theta_i y_i(\theta) + t_i(\theta))\,d\phi_{-i} \quad \text{for all } \phi_{-i} \in \Phi_{-i}.$$

Combining these inequalities implies that, for every measure $\phi = \phi_i \times \phi_{-i} \in \Delta(\Theta_i) \times \Phi_{-i}$,

$$\int_{\Theta_i} \int_{\Theta_{-i}} (\theta_i y_i(\theta) + t_i(\theta))\,d\phi_{-i}\,d\phi_i \geq U_i(\underline{\theta}_i) + \int_{\Theta_i} (1 - F_i(\theta_i))\tilde{y}_i(\theta_i)\,d\theta_i$$

or

$$\int_{\Theta} (\theta_i y_i(\theta) + t_i(\theta))\,d\phi \geq U_i(\underline{\theta}_i) + \int_{\Theta_i} (1 - F_i(\theta_i))\tilde{y}_i(\theta_i)\,d\theta_i. \tag{5}$$

Note that every measure $\phi \in \Phi^*$ is of the form $\phi_i \times \phi_{-i} \in \Delta(\Theta_i) \times \Phi_{-i}$ for each i. Thus, for every $\phi \in \Phi^*$, summing (5) over i yields

$$\sum_i \left(\int_{\Theta} (\theta_i y_i(\theta) + t_i(\theta))\,d\phi \right) \geq \sum_i U_i(\underline{\theta}_i) + \sum_i \left(\int_{\Theta_i} (1 - F_i(\theta_i))\tilde{y}_i(\theta_i)\,d\theta_i \right).$$

Finally, $\sum_i U_i(\underline{\theta}_i) \geq 0$ by IR and $\sum_i \int_{\Theta} t_i(\theta)\,d\phi \leq 0$ by WBB, so this inequality implies (3). □

If truthtelling in the maxmin problem (1) corresponds to a saddle point $(\hat{\theta}_i^*(\theta_i), \phi_{-i}^*(\theta_i))$ (where the agent uses the pure strategy $\hat{\theta}_i^*(\theta_i) = \theta_i$), then letting $y_i^*(\theta_i) = y_i(\theta_i, \phi_{-i}^*(\theta_i))$ be type θ_i's expected allocation under her worst-case belief $\phi_{-i}^*(\theta_i)$, Theorem 4 of Milgrom and Segal (2002) or Theorem 1 of Bodoh-Creed (2012) implies that (4) may be strengthened to

$$U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} y_i^*(s)\,ds \quad \text{for all } \theta_i \in \Theta_i. \tag{6}$$

The same argument as in the proof of Theorem 1 then implies that the necessary condition (3) can be strengthened to

$$\sum_i \left(\int_{\theta \in \Theta} \theta_i y_i(\theta)\,d\phi \right) \geq \sum_i \left(\int_{\theta_i \in \Theta_i} (1 - F_i(\theta_i)) y_i^*(\theta_i)\,d\theta_i \right).$$

Thus, if truthtelling in the maxmin problem corresponds to a saddle point, then a necessary condition for maxmin implementation is that, for every measure $\phi \in \Phi^*$, the expected social surplus under ϕ exceeds the expectation under ϕ of the sum of the agents' subjective information rents (i.e., the information rents under the worst-case beliefs $\phi_{-i}^*(\theta_i)$). Unfortunately, I am not aware of sufficient conditions on the allocation rule

y alone that ensure the existence of such a saddle point.[11] Applying the Debreu–Fan–Glicksberg fixed point theorem to (1), a sufficient condition on the mechanism (y, t) is that y and t are continuous in θ, and $U_i(\hat{\theta}_i, \theta_{-i}; \theta_i)$ is quasiconcave in $\hat{\theta}_i$ and quasiconvex in θ_{-i}.[12]

When the type spaces Θ_i are smoothly path-connected, a well known necessary condition for an *efficient* allocation rule to be Bayesian implementable is that an individually rational VCG mechanism runs an expected surplus (Makowski and Mezzetti 1994, Williams 1999, Krishna and Perry 2000). This follows because the standard envelope characterization of payoffs implies that the interim expected utility of each type in any efficient and Bayesian incentive compatible mechanism is the same as her interim expected utility in a VCG mechanism. However, this result does *not* go through with maxmin incentive compatibility, even if (1) admits a saddle point. This is because the envelope characterization of payoffs with MMIC, (6), depends on types' expected allocations under their worst-case beliefs $\phi^*_{-i}(\theta_i)$, and these beliefs in turn depend on transfers as well as the allocation rule. In particular, distinct efficient and MMIC mechanisms that give the same interim subjective expected utility to the lowest type of each agent need not give the same interim subjective expected utilities to all types, in contrast to the usual payoff equivalence under Bayesian incentive compatibility.[13] Indeed, I show in Section 4 that in the context of bilateral trade, the efficient allocation rule may be maxmin implementable even if all individually rational VCG mechanisms run expected deficits for some measure $\phi \in \Phi^*$.

Conversely, the condition that an individually rational VCG mechanism runs an expected surplus is also *sufficient* for an efficient allocation rule to be Bayesian implementable, because, following Arrow (1979) and d'Aspremont and Gérard-Varet (1979), "lump-sum" transfers that in expectation are constant in θ_i may be used to balance the budget ex post without affecting incentives. This result also does not carry over with maxmin incentive compatibility, as these transfers can affect agents' worst-case beliefs and thereby affect incentives. This issue makes constructing satisfactory MMIC mechanisms challenging, and this paper does not contain positive results on maxmin implementation outside of the bilateral trade context—where, however, a full characterization is provided.

4. Application to bilateral trade

In this section, I show how Theorem 1 can be applied to obtain a full characterization of when efficient bilateral trade is implementable when agents know each other's expected valuation of the good, but know little beyond this.

[11] To clarify, it is easy to provide conditions under which a saddle point in *mixed* strategies is guaranteed to exist (e.g., finiteness of the mechanism). The question here is rather whether there exists a saddle point in which the agent plays a pure strategy, namely the strategy of always reporting her type truthfully.

[12] It should also be noted that if (1) admits a saddle point, then the maximizing and minimizing operators in (1) commute, so that maxmin IC coincides with "minmax" IC. In this case, agents may equivalently be viewed as pessimistic Bayesians rather than as worst-case optimizers.

[13] This point was already noted by Bodoh-Creed (2012).

Formally, a seller s can provide a good at cost $c \in [0, 1]$, and a buyer b values the good at $v \in [0, 1]$.[14] Given a realized cost c and value v, let $y(c, v) \in \{0, 1\}$ and $(t_s(c, v), t_b(v, c))$ denote the resulting allocation (*no trade* or *trade*) and transfers. Thus, efficiency requires that $y(c, v) = 0$ if $c > v$ and $y(c, v) = 1$ if $c < v$.

Note that in the notation of the previous section, $\theta_b = v$ while $\theta_s = -c$. In what follows, I therefore use the notation θ_i to stand for either v (if $i = b$) or $-c$ (if $i = s$).

A key assumption is that the agents' average valuations are known, in that every measure $\phi_i \in \Phi_i$ satisfies $E^{\phi_i}[\theta_i] = \theta_i^*$ for some $\theta_i^* \in (\underline{\theta}_i, \bar{\theta}_i)$.[15] The results in this section actually require only the weaker assumption that $E^{\phi_i}[\theta_i] \geq \theta_i^*$ for all $\phi_i \in \Phi_i$; the intuition is that, with maxmin agents, only bounds on how bad an agent's belief can be are binding.[16] However, Section 5.2 shows that the equality assumption is appropriate if agents have a unique common prior with mean (c^*, v^*) at an ex ante stage and may acquire additional information prior to entering the mechanism. I therefore adopt the equality assumption for consistency with this interpretation.

Two special kinds of distributions ϕ_i will play an important role in the analysis. Let δ_{c^*} be the Dirac measure on c^*, so that $\delta_{c^*} \in \Phi_s$ corresponds to the possibility that the seller's cost is c^* for sure.[17] Let $\delta^s_{c^l, c^h}$ be the two-point measure on c^l and c^h satisfying $E^{\delta^s_{c^l, c^h}}[c] = c^*$; that is, $\delta^s_{c^l, c^h}$ is given by $c = c^l$ with probability $(c^h - c^*)/(c^h - c^l)$ and $c = c^h$ with probability $(c^* - c^l)/(c^h - c^l)$.[18] Thus, $\delta^s_{c^l, c^h} \in \Phi_s$ corresponds to the possibility that the seller's cost may take on only value c^l or c^h.[19] The results to follow require $\delta_{c^*} \in \Phi_s$ and $\delta^s_{c^l, c^h} \in \Phi_s$ for certain values of c^l, c^h (and similarly for the buyer, with the symmetric notation). This richness assumption on the set of the agents' possible beliefs imposes a kind of lower bound on their degree of ambiguity aversion.

It may appear that allowing these Dirac measures introduces an asymmetry with the classical Myerson–Satterthwaite setting, where types are traditionally assumed to be distributed with positive density over intervals. However, all that is required for the Myerson–Satterthwaite theorem to hold is that (i) all types in an interval are "possible," in that incentive compatibility is imposed for an interval of types, and (ii) gains from trade exist with probability strictly between 0 and 1. As per (i), I assume that the set of possible types of each agent is the unit interval. In contrast, it is not clear how to

[14]The sets of possible valuations for the seller and buyer are allowed to differ in the Appendix. This extension is somewhat more significant than in standard Bayesian models, as the set of possible valuations affects the set of feasible worst-cast beliefs.

[15]The assumption that θ_i^* lies in the interior of Θ_i is without loss of generality: if $\theta_i^* \in \{\underline{\theta}_i, \bar{\theta}_i\}$, then there would be no uncertainty about agent i's value, and efficient trade could always be implemented with a VCG mechanism.

[16]Without a bound on how bad beliefs can be, De Castro and Yannelis' (2010) theorem shows that efficient trade is always implementable.

[17]In the information acquisition interpretation of Section 5.2, $\delta_{c^*} \in \Phi_s$ corresponds to the possibility that the seller may acquire no new information about her cost before entering the mechanism.

[18]I omit the superscript on $\delta^s_{c^l, c^h}$ when it is unambiguous.

[19]In the information acquisition interpretation of Section 5.2, $\delta^s_{c^l, c^h} \in \Phi_s$ corresponds to the possibility that the seller may observe a binary signal of her cost before entering the mechanism, where the "good" signal lowers her expected cost to c^l and the "bad" signal raises her expected cost to c^h.

formulate an appropriate analogue of (ii) in the absence of a unique common prior. However, it will turn out that in the $\alpha_i(\theta_i)$ double auction defined below, every type's worst-case belief assigns positive probability to the event that strict gains from trade fail to exist, and efficient trade may be implementable even if the intervals of types that assign positive probability to the event that strict gains from trade do exist overlap.

The characterization result is the following.

Theorem 2. *Assume that $\delta_{0,c} \in \Phi_s$ for all $c \in [c^*, 1]$ and $\delta_{v,1} \in \Phi_b$ for all $v \in [0, v^*]$.*[20] *Then efficient trade is implementable if and only if*

$$\frac{c^*}{1-c^*}\log\left(1+\frac{1-c^*}{c^*}\right)+\frac{1-v^*}{v^*}\log\left(1+\frac{v^*}{1-v^*}\right)\geq 1. \tag{$*$}$$

Theorem 2 shows that, under mild restrictions, efficient bilateral trade between maxmin agents is possible if and only if condition ($*$) holds. In other words, the Myerson–Satterthwaite impossibility result holds with maxmin agents if and only if condition ($*$) fails.

As will become clear, condition ($*$) says precisely that, in the $\alpha_i(\theta_i)$ double auction defined below, $\alpha_s(c)+\alpha_b(v)\leq 1$ for all $c, v \in [0, 1]$. To understand the economic content of condition ($*$), I discuss three aspects of the condition. First, what does condition ($*$) imply for comparative statics and other economic results? Second, where does condition ($*$) come from? And, third, why is condition ($*$) a necessary and sufficient condition for implementation, while Theorem 1 only gives a necessary condition?

To see the implications of condition ($*$), first note that each term in the sum takes the form $(1/x)\log(1+x)$, which is decreasing in x. In particular, decreasing c^* or increasing v^* makes condition ($*$) harder to satisfy. A rough intuition for this comparative static is that improving agent i's average value makes agent j more confident that he will trade, which makes shading his report to get a better transfer more tempting. For example, as $c^* \to 1$ or $v^* \to 0$, the bound on how bad agents' beliefs can be vanishes, and efficient trade is always implementable as in De Castro and Yannelis (2010); conversely, as $c^* \to 0$ or $v^* \to 1$, agents become certain that they will trade, and the temptation to shade their reports becomes irresistible.

Another observation is that condition ($*$) always holds when $c^* \geq v^*$; that is, when the average types of each agent do not have strict gains from trade with each other (e.g., this is why the curve in Figure 1 lies above the 45° line). This follows because, using the inequality $\log 1+x \geq x/(1+x)$, the left-hand side of condition ($*$) is at least $1+c^*-v^*$. This is consistent with Proposition 1 below, which shows that efficient trade is implementable with reference rules when $c^* \geq v^*$. In particular, the parameters for which efficient trade is implementable with general mechanisms but not with reference rules are precisely those that satisfy condition ($*$) but would violate condition ($*$) if the $\log 1+x$ terms were approximated by $x/(1+x)$. This gives one measure of how restrictive reference rules are.

[20] Note that $\delta_{0,c^*} = \delta_{c^*}$, so we have $\delta_{c^*} \in \Phi_s$, and similarly for the buyer.

To see (heuristically) where condition (∗) comes from, suppose that the worst-case belief of a type v buyer who reports type $\hat{v} \geq c^*$ is $\delta^s_{0,\hat{v}}$, which is easily seen to be the belief that minimizes the probability that strict gains from trade exist (i.e., that $c < \hat{v}$) among beliefs ϕ_s with $E^{\phi_s}[c] = c^*$. Suppose also that the mechanism is ex post individually rational. Then

$$U_b(\hat{v}, \delta^s_{0,\hat{v}}; \theta_i) = \frac{\hat{v} - c^*}{\hat{v}}(v + t_b(\hat{v}, 0)) + \frac{c^*}{\hat{v}}(0),$$

where $(\hat{v} - c^*)/\hat{v}$ is the probability of trade (i.e., the probability that $c = 0$ under $\delta^s_{0,\hat{v}}$), and $v + t_b(0, \hat{v})$ is type v's payoff in the event that trade occurs. Assuming that t_b is differentiable, the first-order condition for truthtelling to be optimal is

$$\frac{\partial}{\partial v} t_b(v, 0) = -\frac{c^*}{v(v - c^*)}(v + t_b(v, 0)).$$

This first-order condition captures the trade-off discussed in the Introduction: shading down one's report yields a first-order loss in the probability of trade (i.e., in the probability that $c = 0$), which must be offset by a first-order improvement in the transfer (i.e., in $t_b(v, 0)$).[21]

Solving this differential equation for $t_b(v, 0)$ yields

$$t_b(v, 0) = \frac{v}{v - c^*}(k - c^* \log v),$$

where k is a constant of integration. The constant that keeps transfers bounded as $v \to c^*$ is $k = c^* \log c^*$, which gives

$$t_b(v, 0) = \frac{vc^*}{v - c^*} \log \frac{c^*}{v}.$$

This may be rewritten as

$$t_b(v, 0) = \alpha_b(v)(v - 0) - v,$$

where

$$\alpha_b(v) = 1 - \frac{c^*}{v - c^*} \log \frac{v}{c^*}.$$

The symmetric argument for the seller gives

$$t_s(c, 1) = \alpha_s(c)(1 - c) + c,$$

where

$$\alpha_s(c) = 1 - \frac{1 - v^*}{v^* - c} \log \frac{1 - c}{1 - v^*}.$$

[21] In contrast, for Bayesian agents, shading one's report in a double auction leads to only a second-order loss in foregone gains from trade. This is why double auctions are not incentive compatible for Bayesians.

Now, letting

$$t_b(v,c) = \alpha_b(v)(v-c) - v$$
$$t_s(c,v) = \alpha_s(c)(v-c) + c$$

for all $c < v$, so that the resulting mechanism is an $\alpha_i(\theta_i)$ double auction as described in the Introduction, it may be verified that weak budget balance holds for all (c,v) if and only if it holds for $(c=0, v=1)$ (and it also may be verified that $\delta^s_{0,\hat{v}}$ is indeed a worst-case belief). Therefore, efficient trade is implementable if and only if $t_b(1,0) + t_s(0,1) \leq 0$ or, equivalently, $\alpha_b(1) + \alpha_s(0) \leq 1$. This is precisely condition $(*)$. In other words, condition $(*)$ says that the shares of the social surplus that must be left to the highest types in the $\alpha_i(\theta_i)$ double auction sum to less than 1.

Finally, why does the sufficient condition for implementability that $\alpha_b(1) + \alpha_s(0) \leq 1$ match the necessary condition from Theorem 1? Recall that the necessary condition is that expected social surplus exceeds (a lower bound on) expected information rents (i.e., (3) holds) for any distribution $\phi_s \times \phi_b \in \Phi_s \times \Phi_b$. A first observation is that it suffices to compare the social surplus and information rents under the critical distribution $\delta^s_{0,1} \times \delta^b_{0,1}$, as this distribution may be shown to minimize the difference between the left- and right-hand sides of (3). Note that the expected social surplus under $\delta^s_{0,1} \times \delta^b_{0,1}$ equals $(1-c^*)v^*(1)$, as under $\delta^s_{0,1} \times \delta^b_{0,1}$ there are strict gains from trade only if $c=0$ and $v=1$, which occurs with probability $(1-c^*)v^*$. On the other hand, the expectation of (the lower bound on) the buyer's information rent under $\delta^b_{0,1}$ equals

$$v^* \int_0^1 \tilde{y}_b(v)\, dv + \underbrace{(1-v^*) \int_0^0 \tilde{y}_b(v)\, dv}_{=0},$$

which may be shown to equal $(1-c^*)v^*\alpha_b(1)$. The explanation for the appearance of the $\alpha_b(1)$ term here is that this is the fraction of the social surplus that must be left to a type $v=1$ buyer in an MMIC mechanism when a type v buyer's subjective expected allocation is $\tilde{y}_b(v)$ (in particular, the bound on an agent's subjective information rent given by integrating $\tilde{y}_i(\theta_i)$ is tight in the current setting). Symmetrically, the expectation of the seller's information rent under $\delta^s_{0,1}$ equals $(1-c^*)v^*\alpha_s(0)$, and therefore the necessary condition from Theorem 1 reduces to

$$(1-c^*)v^* \geq (1-c^*)v^*(\alpha_b(1) + \alpha_s(0))$$

or $\alpha_b(1) + \alpha_s(0) \leq 1$.

The approach taken to constructing the $\alpha_i(\theta_i)$ double auction is quite different from standard approaches in Bayesian mechanism design. In particular, the approach here is to posit a type v buyer's worst-case belief to be $\delta^s_{0,v}$ (the belief that minimizes the probability that strict gains from trade exist), to solve a differential equation coming from incentive compatibility for $t_b(v,0)$, which gives the formula for $\alpha_b(v)$, and then to verify that $\delta^s_{0,v}$ is indeed a type v buyer's worst-case belief in the resulting double auction. In contrast, a standard approach might be to use an "off-the-shelf" mechanism, like an AGV mechanism (Arrow 1979, d'Aspremont and Gérard-Varet 1979). However, as argued above, standard arguments for why using such mechanisms is without loss of generality do not apply with maxmin agents; moreover it is not even clear how to define AGV mechanisms in such environments.

A related point is that efficient trade may be implementable even though every individually rational VCG mechanism runs an expected deficit for some measure $\phi \in \Phi^*$, in contrast to the results of Makowski and Mezzetti (1994), Williams (1999), and Krishna and Perry (2000) for Bayesian mechanism design with smoothly path-connected type spaces. For example, this is the case whenever $c^* < v^*$ and condition ($*$) and the assumptions of Theorem 2 hold. To see this, recall that a VCG mechanism is a mechanism where, for all $c, v \in [0, 1]$,

$$t_b(v, c) = -cy(c, v) + h_b(c)$$

for some expected transfer function h_b that depends only on c (and symmetrically for the seller). Note that efficiency and individual rationality of type $v = 0$ imply that $h_b(c^*) \geq 0$ (and symmetrically $h_s(v^*) \geq 0$), as otherwise one would have

$$U_b(0) \leq U_b(0, \delta_{c^*}; 0) = 0 + h_b(c^*) < 0.$$

Hence, the expected deficit of such a mechanism under the measure $\delta_{c^*} \times \delta_{v^*} \in \Phi^*$ equals

$$t_b(v^*, c^*) + t_s(c^*, v^*) = (v^* - c^*)(1) + h_b(c^*) + h_s(v^*) > 0.$$

5. Further results on bilateral trade

This sections presents additional results on bilateral trading with maxmin agents. Section 5.1 characterizes when efficient trade is possible with *reference rules*, a particularly simple class of mechanisms. Section 5.2 describes how the assumption that agents know each other's expected valuation may be interpreted in terms of information acquisition. Section 5.3 discusses the role of Dirac measures in these results, and proposes slight modifications to the definition of the $\alpha_i(\theta_i)$ double auction and reference rule that ensure that these mechanisms are robust to eliminating weakly dominated strategies.

5.1 *Efficient trade with reference rules*

A common justification for introducing concerns about robustness into mechanism design is that these considerations may argue for the use of simpler or otherwise more intuitively appealing mechanisms. The $\alpha_i(\theta_i)$ double auction introduced in the previous section is simple in some ways, but it does involve a carefully chosen transfer rule. In this section, I point out that efficient trade can also be implemented in an extremely simple class of mechanisms—which I call *reference rules*—in the case where the average types of the two agents do not have gains from trade with each other (i.e., when $c^* \geq v^*$).

Reference rules also have the advantage of satisfying strong rather than weak budget balance.[22]

I define a reference rule as follows.

Definition 1. In the bilateral trade setting, a mechanism (y, t) is a *reference rule* if

$$y(c, v) = \begin{cases} 1 & \text{if } c \leq v \\ 0 & \text{if } c > v \end{cases}$$

and there exists a price $p^* \in [0, 1]$ such that

$$t_s(c, v) = -t_b(v, c) = \begin{cases} p^* & \text{if } c \leq p^* \leq v \\ c & \text{if } p^* < c \leq v \\ v & \text{if } c \leq v < p^* \\ 0 & \text{if } c > v. \end{cases}$$

With a reference rule, the agents trade at a *reference price* p^* if they are both willing to do so; otherwise they trade at the reservation price of the agent who is unwilling to trade at the reference price.[23]

Reference rules clearly satisfy EF, (ex post) IR, and SBB, so an MMIC reference rule implements efficient trade. The following result characterizes when MMIC reference rules exist; that is, when efficient trade is implementable with reference rules.

Proposition 1. *Assume that $\delta_{\theta_i^*} \in \Phi_i$ for $i = 1, 2$. Then efficient trade is implementable with reference rules if and only if $c^* \geq v^*$.*

The intuition for why reference rules are incentive compatible when $c^* \geq v^*$ and $p^* \in [v^*, c^*]$ is captured in Figure 2. Observe that every buyer with value $v \leq c^*$ may be certain that no gains from trade exist, as he may believe that the distribution of seller values is the Dirac distribution on c^*. Hence, certainty of no-trade is a worst-case belief for these buyers, and they are therefore willing to reveal their information. In contrast, buyers with value $v > c^*$ do believe that gains from trade exist with positive probability. But it is optimal for these buyers to reveal their values truthfully as well: misreporting some $\hat{v} > c^*$ does not affect the price regardless of the seller's cost (as the price equals c if $c > p^*$ and equals p^* if $c \leq p^*$), and misreporting some $\hat{v} \leq c^*$ again gives payoff 0 in

[22]Another advantage of reference rules is that when $c^* \geq v^*$, they are maxmin incentive compatible in a stronger sense than that of Section 2. First, they remain incentive compatible if agents can hedge ambiguity by randomizing. In addition, they also remain incentive compatible if the order of the maximizing and minimizing operators in (1) is reversed, so that agents are pessimistic Bayesians rather than worst-case optimizers.

[23]The term "reference rule" is taken from Erdil and Klemperer (2010), who recommend the use of such mechanisms in multi-unit auctions. They highlight that reference rules perform well in terms of agents' "local incentives to deviate," a different criterion from what I consider here. Reference rules also bear some resemblance to the "downward flexible price mechanism" of Börgers and Smith (2012). Their mechanism starts with a fixed price p^* which the seller may then lower to any $p' < p^*$, whereupon the parties decide whether to trade at price p'.

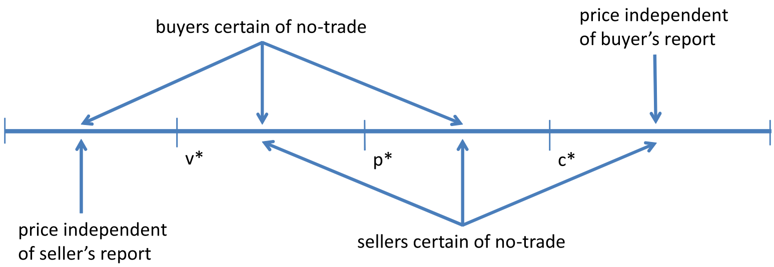

FIGURE 2. Reference rules with $p^* \in [v^*, c^*]$ are incentive compatible.

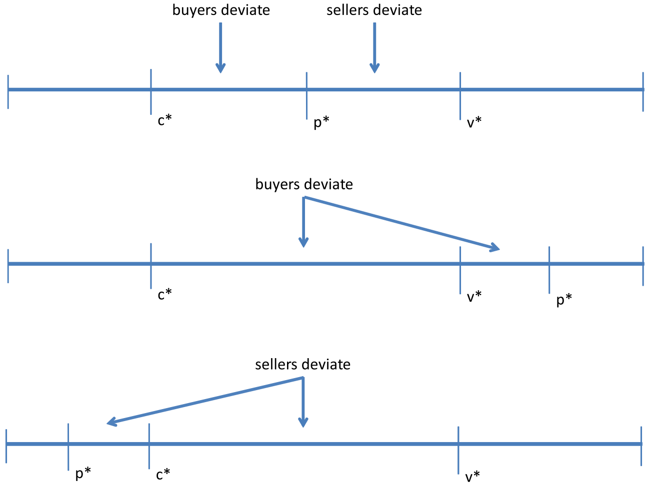

FIGURE 3. When $c^* < v^*$, no reference rule is incentive compatible.

the worst case (as certainty that the seller's value equals c^* would again be a worst-case belief). Therefore, truthtelling is optimal for every buyer type. The argument for sellers is symmetric.

Conversely, Figure 3 indicates why reference rules are not incentive compatible when $c^* < v^*$. Suppose the reference price p^* is greater than c^*. Consider a buyer with value $v \in (c^*, p^*)$. If he reports his value truthfully, then whenever he trades under the reference rule he does so at price v, which gives him payoff 0. Suppose he instead shades his report down to some $\hat{v} \in (c^*, v)$. Then whenever he trades the price is $\hat{v}$—which gives him a positive payoff—and, in addition, he expects to trade with positive probability (since $\hat{v} > c^*$). Hence, he will shade down. The same argument shows that in any reference rule a seller with $c \in (p^*, v^*)$ shades up. Figure 3 shows that a consequence of this argument is that a reference rule cannot be MMIC for both agents at once when $c^* < v^*$, regardless of where the reference price p^* is set.

5.2 *Information acquisition interpretation*

The assumption that agents know the mean and bounds on the support of the distribution of each other's value emerges naturally when agents share a unique common prior at an ex ante stage but are uncertain about the information acquisition technology that the other can access prior to entering the mechanism. This section provides the details of this argument.

Consider the following extension of the model. Each agent i's ex post utility is $\tilde{\theta}_i y + t_i$, where $\tilde{\theta}_i \in \mathbb{R}$ is her realized ex post value. (In the bilateral trade application, the buyer's ex post value is $\tilde{v} = \tilde{\theta}_b$ and the seller's ex post cost is $\tilde{c} = -\tilde{\theta}_s$.) There is an ex ante stage at which the agents' beliefs about the ex post values $(\tilde{\theta}_1, \tilde{\theta}_2)$ are given by a (unique) common product measure $\tilde{\phi}$ on $[\underline{\theta}_1, \bar{\theta}_1] \times [\underline{\theta}_2, \bar{\theta}_2]$ with mean (θ_1^*, θ_2^*) (the common prior). For each agent i, there is a set of possible signaling functions ("experiments") $\mathcal{S}_i$, where a signaling function $\Sigma_i \in \mathcal{S}_i$ is a map from Θ_i to an arbitrary message set M_i, and is thus informative of agent i's own ex post value only. Each agent i knows her own signaling function Σ_i, but is completely uncertain about her opponent's, knowing only that it lies in the set $\mathcal{S}_j$. Agent i's interim value, θ_i—which corresponds to her type in the main model—is her posterior expectation of $\tilde{\theta}_i$ after observing the outcome of her experiment. That is, after observing outcome m_i, agent i's valuation for the good is given by

$$\theta_i \equiv E^{\tilde{\phi}_i}[\tilde{\theta}_i | \Sigma_i(\tilde{\theta}_i) = m_i]. \tag{7}$$

Note that the issue of updating "ambiguous beliefs" does not arise in this model. In particular, the updating in (7) is completely standard. However, the following observation shows that the main model can be interpreted as resulting from each agent's being maxmin about the identity of her opponent's signaling function $\Sigma_j \in \mathcal{S}_j$ at the interim stage (i.e., after she observes her own signal).

Remark 1. If a measure ϕ_i is the distribution of $\theta_i = E^{\tilde{\phi}_i}[\tilde{\theta}_i | \Sigma_i(\tilde{\theta}_i) = m_i]$ under $\tilde{\phi}_i$ for some experiment Σ_i, then $E^{\phi_i}[\theta_i] = \theta_i^*$ and $\operatorname{supp} \phi_i \subseteq \Theta_i$ (where $\operatorname{supp} \phi_i$ denotes the support of ϕ_i).

The fact that $E^{\phi_i}[\theta_i] = \theta_i^*$ is the law of iterated expectation. The fact that $\operatorname{supp} \phi_i \subseteq \Theta_i$ follows because $\tilde{\theta}_i \in [\underline{\theta}_i, \bar{\theta}_i]$ with probability 1 under $\tilde{\phi}_i$. Thus, assuming that agent j finds possible a particular set of measures ϕ_i satisfying $E^{\phi_i}[\theta_i] = \theta_i^*$ and $\operatorname{supp} \phi_i \subseteq \Theta_i$ amounts to assuming that $\mathcal{S}_i$ is a particular subset of the set of all functions $\Theta_i \to M_i$.[24] With this interpretation, the assumption that $\delta_{\theta_i^*} \in \Phi_i$ means that agent j finds it possible that agent i acquires no information about her value before entering the mechanism (beyond the common prior), while the assumption that $\delta^i_{\theta_i^l, \theta_i^h} \in \Phi_i$ means that agent j finds it possible that agent i observes a binary signal of her value, where the bad realization lowers her expectation of $\tilde{\theta}_i$ to θ_i^l and the good realization raises her expectation of $\tilde{\theta}_i$ to θ_i^h.

[24]More precisely, the set of possible interim measures ϕ_i is jointly determined by $\mathcal{S}_i$ and the prior $\tilde{\phi}$. For example, *every* measure ϕ_i such that $E^{\phi_i}[\theta_i] = \theta_i^*$ and $\operatorname{supp} \phi_i \subseteq \Theta_i$ is the distribution of θ_i for some experiment if and only if the prior puts probability 1 on agent i's ex post value being either $\underline{\theta}_i$ or $\bar{\theta}_i$ (see, for example, Theorem 1 of Shmaya and Yariv 2009 or Proposition 1 of Kamenica and Gentzkow 2011).

5.3 *The role of Dirac measures*

In both the $\alpha_i(\theta_i)$ double auction and the reference rule, an agent's worst-case belief is a Dirac (or "two-point") measure. Thus, in these mechanisms, a maxmin agent effectively ignores the outcome that results when her opponent's type takes on all but one or two values. This section argues that this feature is not essential for the results.

A first observation is that excluding Dirac measures per se has no effect on the results if the agents' sets of possible beliefs are sufficiently rich. In particular, if the Dirac measures referenced in the statements of Theorem 2 and Proposition 1 are not contained in Φ_i itself but are contained in its closure, then these results go through as written. For instance, this would be the case if Φ_i consists of all measures on Θ_i with mean θ_i^* that are absolutely continuous with respect to Lebesgue measure. The proof is simply that, with the max inf formulation of (1), excluding accumulation points of Φ_i does not affect agents' utility from any report under any mechanism where $U_j(\hat{\theta}_j, \theta_i; \theta_j)$ is everywhere left or right continuous in θ_i, and both the $\alpha_i(\theta_i)$ double auction and the reference rule satisfy this property.

Nonetheless, one might object that if Dirac measures are viewed as a limiting case in this way, an agent should still not completely discount the possibility that her opponent's type could take on other values. A natural way to formalize this concern is to strengthen the definition of (1) to require that truthtelling is not only maxmin optimal, but also not weakly dominated. I now show that Theorem 2 and Proposition 1 are both robust to this modification, although the mechanisms involved need to be changed slightly.

In the construction in the proof of Theorem 2, $\alpha_b(v) = 0$ if $v < c^*$, so a buyer with value $v < c^*$ gets payoff 0 against every seller type from truthtelling, but gets a positive payoff against some types (and payoff 0 against the others) from shading her report down (and similarly for sellers with $c > v^*$). However, the specification of $\alpha_b(v)$ for types $v < c^*$ can be altered without affecting the desirable properties of the $\alpha_i(\theta_i)$ double auction, as the following result shows. The intuition is that the $\alpha_i(\theta_i)$ double auction runs a strict ex post surplus whenever $v < c^*$ (recall that the surplus is smallest when $v = 1$ and $c = 0$), so some of this surplus can be returned to the buyer without violating budget balance.

Proposition 2. *Theorem 2 continues to hold when the definition of MMIC is strengthened to require that truthtelling is not weakly dominated for any type.*

A similar modification of the reference rule ensures the truthtelling is not weakly dominated: when $v < c^*$, change $t_b(v, c)$ from $-v$ to $-((1 - \varepsilon)v + \varepsilon c)$. However, since reference rules are strongly budget balanced, this modification violates budget balance unless $t_s(c, v)$ is also changed from v to $(1 - \varepsilon)v + \varepsilon c$. This change can in turn lead to a violation of MMIC for the seller. Nonetheless, it turns out that MMIC is preserved if ε is not too large and Φ satisfies the assumptions of Theorem 2.

Proposition 3. *Assume that* $c^* > v^*$ *and that* $\delta_{0,c} \in \Phi_s$ *for all* $c \in [c^*, 1]$ *and* $\delta_{v,1} \in \Phi_b$ *for all* $v \in [0, v^*]$. *Then, for every* $p^* \in (v^*, c^*)$, *the* ε*-modified reference rule given by*

$$t_s(c, v) = -t_b(v, c) = \begin{cases} p^* & \text{if } c \le p^* \le v \\ (1-\varepsilon)c + \varepsilon v & \text{if } p^* < c \le v \\ (1-\varepsilon)v + \varepsilon c & \text{if } c \le v < p^* \\ 0 & \text{if } c > v \end{cases}$$

satisfies EF, IR, and SBB, and satisfies MMIC for all $\varepsilon \in (0, p^*)$. *In addition, under such a mechanism truthtelling is not weakly dominated for any type.*

6. Conclusion

This paper contributes to the study of mechanism design where agents follow robust decision rules, in particular where agents are maxmin expected utility maximizers. I establish two main results. First, I give a general necessary condition for a social choice rule to be implementable, which generalizes the well known condition from Bayesian mechanism design that expected social surplus must exceed expected information rents. This condition involves both a modification of the usual envelope characterization of payoffs and a connection between agents' maxmin expected utilities and the objective expected social surplus under a common prior. Second, I apply this result to give a complete characterization of when efficient bilateral trade is possible, when agents know little beyond each other's expected valuation of the good (which is the information structure that results when agents are maxmin about how one's opponent may acquire information before participating in the mechanism). Somewhat surprisingly, the Myerson–Satterthwaite impossible result sometimes continues to hold with maxmin agents, despite the lack of a unique common prior or independent types. When instead efficient trade is possible, it is implementable with a relatively simple double auction format, the $\alpha_i(\theta_i)$ double auction. Sometimes, it is also implementable with extremely simple reference rules.

A clear direction for future work is investigating positive implementation results beyond the bilateral trade context of two agents and two social alternatives. I have argued that standard mechanisms may fail to have desirable properties with maxmin agents, and in general it is not immediately clear how to generalize the mechanisms I construct in this paper (the $\alpha_i(\theta_i)$ double auction and the reference rule) beyond the bilateral trade case. However, one important setting where a relatively straightforward generalization does exist is the multilateral public good provision problem, where n agents must decide whether to provide a public good at cost C (Mailath and Postlewaite 1990). In this case, the (correlated) belief of a type θ_i agent that minimizes the probability that the good is provided is a two-point distribution, under which either all of her opponents' valuations take on their highest possible values or these valuations sum to $C - \theta_i$ (so strict gains from trade barely fail to exist). This characterization of worst-case beliefs can then be exploited to develop a maxmin incentive compatible trading mechanism, generalizing the $\alpha_i(\theta_i)$ double auction. However, note that these worst-case beliefs involve correlation, while my necessary condition for implementation involves independent beliefs.

This suggests that developing an exact characterization of when efficient public good provision is possible with more than two agents would require a significant extension of the analysis of the bilateral case.

More broadly, it also seems important to consider models of robust agent behavior beyond the maxmin expected utility model. Mechanism design with ambiguity-averse but non-MMEU agents is left for future research, as is mechanism design under other models of robust agent behavior such as minmax regret (Linhart and Radner 1989, Bergemann and Schlag 2008, 2011). The integration of models of robust agent behavior in mechanisms and models of robustness concerns on the part of the mechanism designer (Bergemann and Morris 2005, Chung and Ely 2007) must also await future research.

Appendix: Omitted proofs

Proof of Theorem 2

For the proof of Theorem 2, it is convenient to return to the notation of Section 2, which treats the buyer and seller symmetrically. We also allow for arbitrary type spaces Θ_1 and Θ_2, assuming only that the most favorable type of each agent has gains from trade with the average type of the other agent, while the least favorable type of each agent does not: that is, $\bar{\theta}_i + \theta_j^* > 0 \geq \underline{\theta}_i + \theta_j^*$ for $i = 1, 2$.[25] This assumption is clearly satisfied in the case in the text, where $\bar{\theta}_b + \underline{\theta}_s = \underline{\theta}_b + \bar{\theta}_s = 0$. Noting that condition $(**)$ below generalizes condition $(*)$, the following result generalizes Theorem 2.

Theorem 3. *Assume that $\bar{\theta}_i + \theta_j^* > 0 \geq \underline{\theta}_i + \theta_j^*$ and that $\delta_{\theta_i, \bar{\theta}_i} \in \Phi_i$ for all $\theta_i \in [\underline{\theta}_i, \theta_i^*]$, for $i = 1, 2$. Then efficient trade is implementable if and only if*

$$\begin{aligned}&\left(\frac{\bar{\theta}_1 + \min\{\bar{\theta}_2, -\underline{\theta}_1\}}{\bar{\theta}_1 + \bar{\theta}_2}\right)\left(\frac{\bar{\theta}_1 - \theta_1^*}{\theta_1^* + \min\{\bar{\theta}_2, -\underline{\theta}_1\}}\right)\log\left(1 + \frac{\theta_1^* + \min\{\bar{\theta}_2, -\underline{\theta}_1\}}{\bar{\theta}_1 - \theta_1^*}\right) \\ &\quad + \left(\frac{\bar{\theta}_2 + \min\{\bar{\theta}_1, -\underline{\theta}_2\}}{\bar{\theta}_1 + \bar{\theta}_2}\right)\left(\frac{\bar{\theta}_2 - \theta_2^*}{\theta_2^* + \min\{\bar{\theta}_1, -\underline{\theta}_2\}}\right)\log\left(1 + \frac{\theta_2^* + \min\{\bar{\theta}_1, -\underline{\theta}_2\}}{\bar{\theta}_2 - \theta_2^*}\right) \geq 1.\end{aligned} \tag{$**$}$$

Proof. *Necessity.* By Theorem 1, efficient trade is implementable only if, for all $\phi \in \Phi_1 \times \Phi_2$,

$$\sum_i \left(\int_{\theta \in \Theta} \theta_i y_i(\theta)\, d\phi\right) - \sum_i \left(\int_{\theta_i \in \Theta_i} (1 - F_i(\theta_i))\tilde{y}_i(\theta_i)\, d\theta_i\right) \geq 0 \tag{8}$$

for some allocation rule y_i satisfying

$$y_i(\theta) = \begin{cases} 1 & \text{if } \theta_i + \theta_j > 0 \\ 0 & \text{if } \theta_i + \theta_j < 0. \end{cases}$$

[25] An earlier version of the paper shows that if $\bar{\theta}_i + \theta_j^* \leq 0$ and $\delta_{\theta_j^*} \in \Phi_j$ for some $i \in \{1, 2\}$, then efficient trade is always implementable (with strong budget balance).

Note that, for any such allocation rule y_i,

$$\tilde{y}_i(\theta_i) = \begin{cases} 0 & \text{if } \theta_i \le -\theta_j^* \\ \frac{\theta_j^*+\theta_i}{\bar{\theta}_j+\theta_i} & \text{if } \theta_i \in (-\theta_j^*, -\underline{\theta}_j) \\ 1 & \text{if } \theta_i > -\underline{\theta}_j. \end{cases}$$

This is immediate for the $\theta_i \le -\theta_j^*$ and $\theta_i > -\underline{\theta}_j$ cases, and follows by Chebyshev's inequality in the $\theta_i \in (-\theta_j^*, -\underline{\theta}_j)$ case.[26,27]

Let $\phi_i = \delta_{\max\{\underline{\theta}_i, -\bar{\theta}_j\}, \bar{\theta}_i}$ (which is assumed to be an element of Φ_i, as $-\bar{\theta}_j < \theta_i^*$) for $i = 1, 2$, and let $\phi = \phi_1 \times \phi_2$. Let $\beta_i = (\theta_i^* + \min\{\bar{\theta}_j, -\underline{\theta}_i\})/(\bar{\theta}_i + \min\{\bar{\theta}_j, -\underline{\theta}_i\})$, which is the probability that $\theta_i = \bar{\theta}_i$ under $\delta_{\max\{\underline{\theta}_i, -\bar{\theta}_j\}, \bar{\theta}_i}$. Observe that

$$\sum_i \left(\int_{\theta \in \Theta} \theta_i y_i(\theta)\, d\phi \right) = (\bar{\theta}_i + \bar{\theta}_j)\beta_i\beta_j + \max\{\bar{\theta}_i + \underline{\theta}_j, 0\}\beta_i(1-\beta_j) + \max\{\bar{\theta}_j + \underline{\theta}_i, 0\}\beta_j(1-\beta_i)$$

and, using the assumption that $\underline{\theta}_i + \theta_j^* \le 0$,

$$\begin{aligned} \int_{\theta_i \in \Theta_i} (1 - F_i(\theta))\tilde{y}_i(\theta_i)\, d\theta_i &= \int_{\max\{\underline{\theta}_i, -\theta_j^*\}}^{\min\{\bar{\theta}_i, -\underline{\theta}_j\}} \beta_i \left(\frac{\theta_j^* + \theta_i}{\bar{\theta}_j + \theta_i} \right) d\theta_i + \int_{\min\{\bar{\theta}_i, -\underline{\theta}_j\}}^{\bar{\theta}_i} \beta_i\, d\theta_i \\ &= (\bar{\theta}_i + \theta_j^*)\beta_i - (\bar{\theta}_j - \theta_j^*)\beta_i \log\left(1 + \frac{\theta_j^* + \min\{\bar{\theta}_i, -\underline{\theta}_j\}}{\bar{\theta}_j - \theta_j^*} \right). \end{aligned}$$

Combining these observations and collecting terms, the left-hand side of (8) equals

$$\begin{aligned} \zeta\Bigg[-1 &+ \left(\frac{\bar{\theta}_1 + \min\{\bar{\theta}_2, -\underline{\theta}_1\}}{\bar{\theta}_1 + \bar{\theta}_2} \right)\left(\frac{\bar{\theta}_1 - \theta_1^*}{\theta_1^* + \min\{\bar{\theta}_2, -\underline{\theta}_1\}} \right) \log\left(1 + \frac{\theta_1^* + \min\{\bar{\theta}_2, -\underline{\theta}_1\}}{\bar{\theta}_1 - \theta_1^*} \right) \\ &+ \left(\frac{\bar{\theta}_2 + \min\{\bar{\theta}_1, -\underline{\theta}_2\}}{\bar{\theta}_1 + \bar{\theta}_2} \right)\left(\frac{\bar{\theta}_2 - \theta_2^*}{\theta_2^* + \min\{\bar{\theta}_1, -\underline{\theta}_2\}} \right) \log\left(1 + \frac{\theta_2^* + \min\{\bar{\theta}_1, -\underline{\theta}_2\}}{\bar{\theta}_2 - \theta_2^*} \right) \Bigg], \end{aligned} \tag{9}$$

where $\zeta = \beta_1\beta_2(\bar{\theta}_1 + \bar{\theta}_2) > 0$. The bracketed term in (9) nonnegative if and only if condition $(**)$ holds. Hence, condition $(**)$ is necessary.

Sufficiency. The $\alpha_i(\theta_i)$ double auction is defined by

$$y(\theta_i, \theta_j) = \begin{cases} 1 & \text{if } \theta_i + \theta_j > 0 \\ 0 & \text{if } \theta_i + \theta_j \le 0 \end{cases}$$

$$t_i(\theta_i, \theta_j) = \begin{cases} \alpha_i(\theta_i)\theta_j - (1 - \alpha_i(\theta_i))\min\{\theta_i, -\underline{\theta}_j\} & \text{if } \theta_i + \theta_j > 0 \\ 0 & \text{if } \theta_i + \theta_j \le 0 \end{cases}$$

[26] The form of Chebyshev's inequality I use throughout the paper is, for random variable X with mean x^* and upper bound $\bar{x}$, $\Pr(X \ge x) \ge (x^* - x)/(\bar{x} - x)$. This follows because $x^* \le \Pr(X \ge x)\bar{x} + \Pr(X < x)x$. See, for example, p. 319 of Grimmett and Stirzaker (2001).

[27] As will become clear, the value of $\tilde{y}_i(-\underline{\theta}_j)$ does not matter for the proof.

for $i = 1, 2$, where

$$\alpha_i(\theta_i) = \begin{cases} 1 - \frac{\bar{\theta}_j - \theta_j^*}{\theta_j^* + \min\{\theta_i, -\underline{\theta}_j\}} \log(1 + \frac{\theta_j^* + \min\{\theta_i, -\underline{\theta}_j\}}{\bar{\theta}_j - \theta_j^*}) & \text{if } \theta_i > -\theta_j^* \\ 0 & \text{if } \theta_i \le -\theta_j^*. \end{cases}$$

This mechanism is clearly efficient. I show that it satisfies IR and MMIC, and that it satisfies WBB if and only if condition (∗∗) holds.

Claim 1. *The $\alpha_i(\theta_i)$ double auction satisfies (ex post) IR.*

Proof. If $\theta_i + \theta_j \le 0$, then $U_i(\theta_i, \theta_j; \theta_i) = 0$. If $\theta_i + \theta_j > 0$, then

$$\begin{aligned} U_i(\theta_i, \theta_j; \theta_i) &= \theta_i + \alpha_i(\theta_i)\theta_j - (1 - \alpha_i(\theta_i)) \min\{\theta_i, -\underline{\theta}_j\} \\ &\ge \alpha_i(\theta_i)(\theta_i + \theta_j). \end{aligned}$$

Now $\alpha_i(\theta_i)$ is of the form $1 - (1/x)\log(1 + x)$ for $x > 0$ and $(1/x)\log(1 + x) \in (0, 1)$ for $x > 0$, so $\alpha_i(\theta_i) \in (0, 1)$ for all θ_i. This yields (ex post) IR. ◁

Claim 2. *The $\alpha_i(\theta_i)$ double auction satisfies WBB if and only if condition (∗∗) holds.*

Proof. WBB is trivially satisfied when $\theta_1 + \theta_2 \le 0$, so suppose that $\theta_1 + \theta_2 > 0$.

If $\theta_1 < -\underline{\theta}_2$ and $\theta_2 < -\underline{\theta}_1$,

$$t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) = (\theta_1 + \theta_2)(\alpha_1(\theta_1) + \alpha_2(\theta_2) - 1).$$

Since $\alpha_i(\theta_i)$ is nondecreasing in θ_i (as $(1/x)\log(1+x)$ is decreasing in x), this expression is nonpositive for all θ_1, θ_2 with $\theta_1 + \theta_2 > 0$ if and only if $\alpha_1(\bar{\theta}_1) + \alpha_2(\bar{\theta}_2) \le 1$. Condition (∗∗) implies $\alpha_1(\bar{\theta}_1) + \alpha_2(\bar{\theta}_2) \le 1$ and is equivalent to this inequality when $\bar{\theta}_i \le -\underline{\theta}_j$ for $i = 1, 2$.

If $\theta_1 \ge -\underline{\theta}_2$ and $\theta_2 \ge -\underline{\theta}_1$, then

$$\begin{aligned} t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) &= \alpha_1(\theta_1)\theta_2 + \alpha_2(\theta_2)\theta_1 + (1 - \alpha_1(\theta_1))\underline{\theta}_2 + (1 - \alpha_2(\theta_2))\underline{\theta}_1 \\ &= \theta_1 + \theta_2 - (1 - \alpha_1(\theta_1))(\theta_2 - \underline{\theta}_2) - (1 - \alpha_2(\theta_2))(\theta_1 - \underline{\theta}_1). \end{aligned}$$

This expression is nondecreasing in θ_1 and θ_2 (as $\alpha_i(\theta_i) \in (0, 1)$ is nondecreasing and $\theta_i \ge \underline{\theta}_i$), so it is nonpositive for all θ_1, θ_2 with $\theta_1 + \theta_2 > 0$ if and only if

$$\bar{\theta}_1 + \bar{\theta}_2 - (1 - \alpha_1(\bar{\theta}_1))(\bar{\theta}_2 - \underline{\theta}_2) - (1 - \alpha_2(\bar{\theta}_2))(\bar{\theta}_1 - \underline{\theta}_1) \le 0.$$

Moving the product terms to the right-hand side and dividing by $\bar{\theta}_1 + \bar{\theta}_2$ (which is positive) shows that this inequality is equivalent to condition (∗∗) when $\bar{\theta}_i \ge -\underline{\theta}_j$ for $i = 1, 2$ (which is the case under consideration).

Finally, if $\theta_1 < -\underline{\theta}_2$ and $\theta_2 \ge -\underline{\theta}_1$ (which is the hardest case),

$$\begin{aligned} t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) &= (\alpha_1(\theta_1) + \alpha_2(\theta_2) - 1)\theta_1 + \alpha_1(\theta_1)\theta_2 + (1 - \alpha_2(\theta_2))\underline{\theta}_1 \\ &= (\theta_1 + \theta_2)\left[\alpha_1(\theta_1) - \frac{\theta_1 - \underline{\theta}_1}{\theta_1 + \theta_2}(1 - \alpha_2(\theta_2))\right]. \end{aligned}$$

This expression is nonpositive for all θ_1, θ_2 with $\theta_1 + \theta_2 > 0$ if and only if the bracketed term is nonpositive for all such θ_1, θ_2. This term is increasing in θ_2, so it is nonpositive for all θ_1, θ_2 with $\theta_1 + \theta_2 > 0$ if and only if

$$\alpha_1(\theta_1) - \frac{\theta_1 - \underline{\theta}_1}{\theta_1 + \bar{\theta}_2}(1 - \alpha_2(\bar{\theta}_2)) \leq 0 \tag{10}$$

for all θ_1.[28] If $\theta_1 \leq -\theta_2^*$, then $\alpha_1(\theta_1) = 0$ so (10) holds. Suppose toward a contradiction that (10) fails for some $\theta_1 \in [-\theta_2^*, \min\{\bar{\theta}_1, -\underline{\theta}_2\}]$. Observe first that (10) holds at $\theta_1 = \min\{\bar{\theta}_1, -\underline{\theta}_2\}$: this has already been shown if $\bar{\theta}_1 \geq -\underline{\theta}_2$, and if $\bar{\theta}_1 < -\underline{\theta}_2$, it follows by noting that at $\theta_1 = \bar{\theta}_1$ (10) is equivalent to condition (∗∗) when $\bar{\theta}_1 < -\underline{\theta}_2$ and $\bar{\theta}_2 \geq -\underline{\theta}_1$. Since the left-hand side of (10) is continuous in θ_1 and (10) holds for $\theta_1 = \bar{\theta}_1$ and for all $\theta_1 \geq -\underline{\theta}_2$, (10) fails somewhere on the interval $[-\theta_2^*, \min\{\bar{\theta}_1, -\underline{\theta}_2\}]$ if and only if it fails at a local minimum in $(-\theta_2^*, -\underline{\theta}_2)$. Hence, the argument may be completed by showing that no local minimum in $(-\theta_2^*, -\underline{\theta}_2)$ exists. To see this, note that for $\theta_1 \in (-\theta_2^*, -\underline{\theta}_2)$,

$$\alpha_1'(\theta_1) = \frac{1}{\theta_2^* + \theta_1}\left(\frac{\theta_2^* + \theta_1}{\bar{\theta}_2 + \theta_1} - \alpha_1(\theta_1)\right),$$

and therefore the first-order condition for an extremum is

$$\frac{1}{\theta_2^* + \theta_1}\left(\frac{\theta_2^* + \theta_1}{\bar{\theta}_2 + \theta_1} - \alpha_1(\theta_1)\right) = \frac{\bar{\theta}_2 + \underline{\theta}_1}{(\bar{\theta}_2 + \theta_1)^2}(1 - \alpha_2(\bar{\theta}_2)).$$

In addition, the second derivative of the left-hand side of (10) equals

$$-\frac{(\bar{\theta}_2 - \theta_2^*)(2(\bar{\theta}_2 + \theta_1) + \theta_2^* + \theta_1)}{(\theta_2^* + \theta_1)^2(\bar{\theta}_2 + \theta_1)^2} + 2\frac{1 - \alpha_1(\theta_1)}{(\theta_2^* + \theta_1)^2} - 2\frac{\bar{\theta}_2 + \underline{\theta}_1}{(\bar{\theta}_2 + \theta_1)^3}(1 - \alpha_2(\bar{\theta}_2)).$$

At an extremum, using the first-order condition implies that this equals

$$\begin{aligned}
&-\frac{(\bar{\theta}_2 - \theta_2^*)(2(\bar{\theta}_2 + \theta_1) + \theta_2^* + \theta_1)}{(\theta_2^* + \theta_1)^2(\bar{\theta}_2 + \theta_1)^2} + 2\frac{1 - \alpha_1(\theta_1)}{(\theta_2^* + \theta_1)^2} \\
&\qquad\qquad - 2\frac{1}{(\bar{\theta}_2 + \theta_1)(\theta_2^* + \theta_1)}\left(\frac{\theta_2^* + \theta_1}{\bar{\theta}_2 + \theta_1} - \alpha_1(\theta_1)\right) \\
&= \frac{\bar{\theta}_2 - \theta_2^*}{(\theta_2^* + \theta_1)^2(\bar{\theta}_2 + \theta_1)}\left[-\alpha_1(\theta_1) + \left(\frac{\theta_2^* + \theta_1}{\bar{\theta}_2 + \theta_1} - \alpha_1(\theta_1)\right)\right] \\
&= \frac{\bar{\theta}_2 - \theta_2^*}{(\theta_2^* + \theta_1)^2(\bar{\theta}_2 + \theta_1)}\left[-\alpha_1(\theta_1) + \frac{(\theta_2^* + \theta_1)(\bar{\theta}_2 + \underline{\theta}_1)}{(\bar{\theta}_2 + \theta_1)^2}(1 - \alpha_2(\bar{\theta}_2))\right].
\end{aligned}$$

Next, observe that

$$\frac{\theta_1 - \underline{\theta}_1}{\theta_1 + \bar{\theta}_2} \geq \frac{(\theta_2^* + \theta_1)(\bar{\theta}_2 + \underline{\theta}_1)}{(\bar{\theta}_2 + \theta_1)^2},$$

[28] The "only if" part of this statement follows because the hypothesis that $\theta_2 \geq -\underline{\theta}_1$ implies that $\bar{\theta}_2 \geq -\underline{\theta}_1$, which in turn implies that $\theta_1 + \bar{\theta}_2 \geq 0$ for all θ_1.

which may be seen by cross-multiplying by $(\bar{\theta}_2+\theta_1)^2$ and noting that $\theta_1-\underline{\theta}_1 \geq \theta_2^*+\theta_1$ (as $\underline{\theta}_1+\theta_2^* \leq 0$) and $\bar{\theta}_2+\theta_1 \geq \bar{\theta}_2+\underline{\theta}_1$. Therefore,

$$-\alpha_1(\theta_1)+\frac{\theta_1-\underline{\theta}_1}{\theta_1+\bar{\theta}_2}(1-\alpha_2(\bar{\theta}_2)) \geq -\alpha_1(\theta_1)+\frac{(\bar{\theta}_2+\underline{\theta}_1)(\theta_2^*+\theta_1)}{(\theta_1+\bar{\theta}_2)^2}(1-\alpha_2(\bar{\theta}_2)),$$

so if (10) fails, then the second derivative is nonpositive at any local extremum. That is, any local extremum in $(-\theta_2^*, -\underline{\theta}_2)$ must be a local maximum, so no local minimum in $(-\theta_2^*, -\underline{\theta}_2)$ exists, completing the proof. The argument for $\theta_1 \geq -\underline{\theta}_2$ and $\theta_2 < -\underline{\theta}_1$ is symmetric. $\triangleleft$

Claim 3. *The $\alpha_i(\theta_i)$ double auction satisfies MMIC.*

Proof. Suppose $\theta_i \leq -\theta_j^*$. By IR, $U_i(\theta_i) \geq 0$. By ex post IR and WBB, $t_i(\hat{\theta}_i, \theta_j) \leq \theta_j$ for all $\hat{\theta}_i$, θ_j, and therefore $U_i(\hat{\theta}_i, \theta_j^*; \theta_i) \leq \max\{\theta_i+\theta_j^*, 0\} \leq 0$ for all $\hat{\theta}_i$. Hence, $\delta_{\theta_j^*} \in \Phi_j$ implies that $U_i(\theta_i) \geq U_i(\hat{\theta}_i, \theta_j^*; \theta_i) \geq \inf_{\phi_j \in \Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i)$, which yields MMIC.

For the remainder of the proof, suppose $\theta_i > -\theta_j^*$. I show that no misreport $\hat{\theta}_i$ can be profitable in each of the four cases (i) $\hat{\theta}_i > \theta_i$, (ii) $\hat{\theta}_i \leq -\theta_j^*$, (iii) $\hat{\theta}_i \in [-\underline{\theta}_j, \theta_i)$ (this case is vacuous if $\theta_i \leq -\underline{\theta}_j$), and (iv) $\hat{\theta}_i \in (-\theta_j^*, \min\{\theta_i, -\underline{\theta}_j\})$. These cases cover all possible misreports, so the $\alpha_i(\theta_i)$ double auction satisfies MMIC.

Case (i): $\hat{\theta}_i > \theta_i$. In this case, I claim that $U_i(\theta_i, \theta_j; \theta_i) \geq U_i(\hat{\theta}_i, \theta_j; \theta_i)$ for all θ_j. The key step is the following observation (proven below).

Lemma 2. *In the $\alpha_i(\theta_i)$ double auction, $t_i(\theta_i, \theta_j)$ is nonincreasing in θ_i in the region where $\theta_i+\theta_j > 0$.*

Now, if $\theta_i+\theta_j \leq 0$, then $U_i(\theta_i, \theta_j; \theta_i) = 0$, while ex post IR and WBB imply that $U_i(\hat{\theta}_i, \theta_j; \theta_i) \leq \max\{\theta_i+\theta_j, 0\} \leq 0$. If instead $\theta_i+\theta_j > 0$, then EF and Lemma 2 imply that $U_i(\theta_i, \theta_j; \theta_i) \geq U_i(\hat{\theta}_i, \theta_j; \theta_i)$. The claim follows, and therefore $U_i(\theta_i) \geq \inf_{\phi_j \in \Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i)$.

Case (ii): $\hat{\theta}_i \leq -\theta_j^*$. Here, $U_i(\hat{\theta}_i, \theta_j^*; \theta_i) = 0$. Hence, $\delta_{\theta_j^*} \in \Phi_j$ and IR imply that $U_i(\theta_i) \geq \inf_{\phi_j \in \Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i)$.

Case (iii): $\hat{\theta}_i \in [-\underline{\theta}_j, \theta_i)$. Note that $\alpha_i(\hat{\theta}_i) = \alpha_i(-\underline{\theta}_j)$, so $U_i(\hat{\theta}_i, \theta_j; \theta_i) = \theta_i + \alpha_i(-\underline{\theta}_j)\theta_j - (1-\alpha_i(-\underline{\theta}_j))\underline{\theta}_j$ for all θ_j. Therefore, $U_i(\hat{\theta}_i, \phi_j; \theta_i) = \theta_i + \alpha_i(-\underline{\theta}_j)\theta_j^* - (1-\alpha_i(-\underline{\theta}_j))\underline{\theta}_j$ for all $\phi_j \in \Phi_j$. Similarly, $U_i(\theta_i, \phi_j; \theta_i) = \theta_i + \alpha_i(-\underline{\theta}_j)\theta_j^* - (1-\alpha_i(-\underline{\theta}_j))\underline{\theta}_j$, so $U_i(\theta_i) = \inf_{\phi_j \in \Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i)$.

Case (iv): $\hat{\theta}_i \in (-\theta_j^*, \min\{\theta_i, -\underline{\theta}_j\})$. In this case, I claim that

$$\inf_{\phi_j \in \Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i) = \frac{\theta_j^*+\hat{\theta}_i}{\bar{\theta}_j+\hat{\theta}_i}(\theta_i + t_i(\hat{\theta}_i, \bar{\theta}_j)). \tag{11}$$

(Intuitively, the claim is that $\delta_{\max\{-\hat{\theta}_i, \underline{\theta}_j\}, \bar{\theta}_j}$ is a worst-case belief for an agent of type θ_i who misreports as type $\hat{\theta}_i \in (-\theta_j^*, \theta_i]$.)

To see that (11) is an upper bound on $\inf_{\phi_j\in\Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i)$, observe that $\delta_{-\hat{\theta}_i,\bar{\theta}_j} \in \Phi_j$ and that $U_i(\hat{\theta}_i, \delta_{-\hat{\theta}_i,\bar{\theta}_j}; \theta_i)$ equals (11). To see that (11) is a lower bound on $\inf_{\phi_j\in\Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i)$, first note that

$$\begin{aligned} U_i(\hat{\theta}_i, \phi_j; \theta_i) &= \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)E^{\phi_j}[\theta_i + t_i(\hat{\theta}_i, \theta_j)|\theta_j > -\hat{\theta}_i] + \Pr{}^{\phi_j}(\theta_j \le -\hat{\theta}_i)(0) \\ &= \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)E^{\phi_j}[\theta_i + \alpha_i(\hat{\theta}_i)\theta_j - (1-\alpha_i(\hat{\theta}_i))\hat{\theta}_i|\theta_j > -\hat{\theta}_i] \\ &= \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)(\theta_i - \hat{\theta}_i) + \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)\alpha_i(\hat{\theta}_i)(E^{\phi_j}[\theta_j|\theta_j > -\hat{\theta}_i] + \hat{\theta}_i). \end{aligned} \tag{12}$$

I show that (11) is a lower bound on (12) for all ϕ_j with expectation θ_j^*, and hence for all $\phi_j \in \Phi_j$. To see this, consider the problem of minimizing (12) over ϕ_j with expectation θ_j^* in two steps: first minimize over ϕ_j with a given value of $\Pr^{\phi_j}(\theta_j > -\hat{\theta}_i)$, and then minimize over $\Pr^{\phi_j}(\theta_j > -\hat{\theta}_i)$. For a given value of $\Pr^{\phi_j}(\theta_j > -\hat{\theta}_i)$, (12) is minimized by minimizing $E^{\phi_j}[\theta_j|\theta_j > -\hat{\theta}_i]$ over ϕ_j with expectation θ_j^*. Observe that

$$(1 - \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i))E^{\phi_j}[\theta_j|\theta_j \le -\hat{\theta}_i] + \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)E^{\phi_j}[\theta_j|\theta_j > -\hat{\theta}_i] = \theta_j^*$$

for all ϕ_j with expectation θ_j^*. Noting that $E^{\phi_j}[\theta_j|\theta_j \le -\hat{\theta}_i] \le -\hat{\theta}_i$ and rearranging yields

$$E^{\phi_j}[\theta_j|\theta_j > -\hat{\theta}_i] \ge \frac{1}{\Pr^{\phi_j}(\theta_j > -\hat{\theta}_i)}\big(\theta_j^* + (1 - \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i))\hat{\theta}_i\big).$$

Hence, the minimum of (12) over ϕ_j with expectation θ_j^* and a given value of $\Pr^{\phi_j}(\theta_j > -\hat{\theta}_i)$ equals

$$\begin{aligned} &\Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)(\theta_i - \hat{\theta}_i) \\ &\qquad + \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)\alpha_i(\hat{\theta}_i)\left(\frac{1}{\Pr^{\phi_j}(\theta_j > -\hat{\theta}_i)}\big(\theta_j^* + (1 - \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i))\hat{\theta}_i\big) + \hat{\theta}_i\right) \\ &= \Pr{}^{\phi_j}(\theta_j > -\hat{\theta}_i)(\theta_i - \hat{\theta}_i) + \alpha_i(\hat{\theta}_i)(\theta_j^* + \hat{\theta}_i). \end{aligned}$$

As $\hat{\theta}_i \le \theta_i$, (12) is minimized over ϕ_j with expectation θ_j^* by minimizing $\Pr^{\phi_j}(\theta_j > -\hat{\theta}_i)$, which by Chebyshev's inequality yields

$$\begin{aligned} \frac{\theta_j^* + \hat{\theta}_i}{\bar{\theta}_j + \hat{\theta}_i}(\theta_i - \hat{\theta}_i) + \alpha_i(\hat{\theta}_i)(\theta_j^* + \hat{\theta}_i) &= \frac{\theta_j^* + \hat{\theta}_i}{\bar{\theta}_j + \hat{\theta}_i}(\theta_i - \hat{\theta}_i + \alpha_i(\hat{\theta}_i)(\bar{\theta}_j + \hat{\theta}_i)) \\ &= \frac{\theta_j^* + \hat{\theta}_i}{\bar{\theta}_j + \hat{\theta}_i}(\theta_i + t_i(\hat{\theta}_i, \bar{\theta}_j)). \end{aligned}$$

This gives (11), proving the claim. ◁

Therefore,

$$\sup_{\hat{\theta}_i\in(-\theta_j^*,\min\{\theta_i,-\underline{\theta}_j\})} \inf_{\phi_j\in\Phi_j} U_i(\hat{\theta}_i, \phi_j; \theta_i) = \sup_{\hat{\theta}_i\in(-\theta_j^*,\min\{\theta_i,-\underline{\theta}_j\})} \frac{\theta_j^* + \hat{\theta}_i}{\bar{\theta}_j + \hat{\theta}_i}(\theta_i + t_i(\hat{\theta}_i, \bar{\theta}_j)).$$

To complete the proof, it suffices to show that $((\theta_j^* + \hat{\theta}_i)/(\bar{\theta}_j + \hat{\theta}_i))(\theta_i + t_i(\hat{\theta}_i, \bar{\theta}_j))$ is nondecreasing in $\hat{\theta}_i$ over $(-\theta_j^*, \min\{\theta_i, -\underline{\theta}_j\})$, as $U_i(\hat{\theta}_i, \theta_j; \theta_i)$ is left-continuous in $\hat{\theta}_i$ and the possibility that $-\underline{\theta}_j$ could be a profitable misreport has already been ruled out. This follows, since a straightforward calculation yields

$$\frac{\partial}{\partial \hat{\theta}_i}\left[\frac{\theta_j^* + \hat{\theta}_i}{\bar{\theta}_j + \hat{\theta}_i}(\theta_i + t_i(\hat{\theta}_i, \bar{\theta}_j))\right] = \frac{(\bar{\theta}_j - \theta_j^*)(\theta_i - \hat{\theta}_i)}{(\bar{\theta}_j + \hat{\theta}_i)^2},$$

and this expression is nonnegative because $\theta_i \geq \hat{\theta}_i \geq -\theta_j^* \geq -\bar{\theta}_j$. □

Proof of Lemma 2. The result is immediate when $\theta_i \leq -\theta_j^*$ or $\theta_i \geq -\underline{\theta}_j$, as in both cases $\alpha_i'(\theta_i) = 0$, which immediately implies that $t_i(\theta_i, \theta_j)$ is nonincreasing in θ_i.

If $\theta_i \in (-\theta_j^*, -\underline{\theta}_j)$, then $t_i(\theta_i, \theta_j) = \alpha_i(\theta_i)\theta_j - (1 - \alpha_i(\theta_i))\theta_i$, and therefore

$$\frac{\partial}{\partial \theta_i} t_i(\theta_i, \theta_j) = -(1 - \alpha_i(\theta_i)) + \alpha_i'(\theta_i)(\theta_j + \theta_i).$$

In addition,

$$\alpha_i'(\theta_i) = \frac{1}{\bar{\theta}_j + \theta_i} - \frac{1}{\theta_j^* + \theta_i}\alpha_i(\theta_i),$$

and therefore

$$\begin{aligned}\frac{\partial}{\partial \theta_i} t_i(\theta_i, \theta_j) &= \frac{\theta_j - \theta_j^*}{\theta_j^* + \theta_i}(1 - \alpha_i(\theta_i)) - \frac{\bar{\theta}_j - \theta_j^*}{\theta_j^* + \theta_i}\frac{\theta_j + \theta_i}{\bar{\theta}_j + \theta_i} \\ &= \frac{\bar{\theta}_j - \theta_j^*}{\theta_j^* + \theta_i}\left[\frac{\theta_j - \theta_j^*}{\theta_j^* + \theta_i}\log\left(1 + \frac{\theta_j^* + \theta_i}{\bar{\theta}_j - \theta_j^*}\right) - \frac{\theta_j + \theta_i}{\bar{\theta}_j + \theta_i}\right].\end{aligned}$$

Since $\theta_i > -\theta_j^*$, the sign of $\partial t_i(\theta_i, \theta_j)/\partial \theta_i$ equals the sign of the term in brackets. Using the fact that $(1/x)\log(1 + x) < 1$ for all $x > 0$, this term is less than

$$\frac{\theta_j - \theta_j^*}{\bar{\theta}_j - \theta_j^*} - \frac{\theta_j + \theta_i}{\bar{\theta}_j + \theta_i} = -\frac{(\bar{\theta}_j - \theta_j)(\theta_j^* + \theta_i)}{(\bar{\theta}_j - \theta_j^*)(\bar{\theta}_j + \theta_i)} \leq 0,$$

where the last inequality again uses $\theta_i > -\theta_j^*$. Hence, $t_i(\theta_i, \theta_j)$ is nonincreasing in θ_i. □

Proof of Proposition 1

Sufficiency. When $c^* \geq v^*$, I show that any reference rule with $p^* \in [v^*, c^*]$ satisfies MMIC. I establish MMIC for the buyer; the argument for the seller is symmetric.

First, suppose that $v < c^*$. Observe that $U_b(\hat{v}, c^*; v) \leq 0$ for all $\hat{v}$: this follows because if $\hat{v} < c^*$, then $U_b(\hat{v}, c^*; v) = 0$, while if $\hat{v} \geq c^*$, then $U_b(\hat{v}, c^*; v) = v - c^* < 0$. Hence, for all $\hat{v}$, IR and $\delta_{c^*} \in \Phi_s$ imply that $U_b(v) \geq 0 \geq \inf_{\phi_s \in \Phi_s} U_b(\hat{v}, \phi_s; v)$, which yields MMIC.

Next, suppose that $v \geq c^*$. First, note that misreports of $\hat{v} \leq c^*$ cannot be profitable, because $U_b(v) \geq 0 = U_b(\hat{v}, c^*; v)$ and $\delta_{c^*} \in \Phi_s$. Next, consider misreports of $\hat{v} > c^*$.

If $y(c, v) = y(c, \hat{v})$, then $t_b(v, c) = t_b(\hat{v}, c)$ (as $v, \hat{v} \geq c^* \geq p^*$), and hence $U_b(v, c; v) = U_b(\hat{v}, c; v)$. In addition, if $y(c, v) = 1$ and $y(c, \hat{v}) = 0$, then $U_b(v, c; v) \geq 0 = U(\hat{v}, c; v)$. Finally, if $y(c, v) = 0$ and $y(c, \hat{v}) = 1$, then $U_b(v, c; v) = 0 > v - c \geq U_b(\hat{v}, c; v)$. Hence, $U_b(v, c; v) \geq U_b(\hat{v}, c; v)$ for all $\hat{v}$, so misreports of $\hat{v} > c^*$ cannot be profitable, either. This yields MMIC.

Necessity. When $c^* < v^*$, every reference rule satisfies either $p^* > c^*$ or $p^* < v^*$. Suppose $p^* > c^*$; the argument for the alternative case is symmetric. Fix a value $v \in [c^*, p^*] \cap [0, 1]$. Then $t_b(v, c) = -v$ whenever $c \leq v$. Hence, a buyer with value v gets payoff 0 from truthtelling, and gets a strictly positive payoff from reporting any $v \in (c^*, v)$ (as every belief in Φ_s puts positive probability on seller types with $c \leq c^*$), so MMIC fails.

Proof of Proposition 2

We show that the slightly more general Theorem 3 goes through with this more restrictive definition of MMIC. The notation in what follows is as in the proof of Theorem 3.

The proof of necessity is unchanged. For sufficiency, modify the $\alpha_i(\theta_i)$ double auction constructed in the proof of Theorem 3 by letting $\alpha_i(\theta_i) = \min\{\frac{1}{2}, (\theta_i - \underline{\theta}_i)/(\bar{\theta}_j - \theta_j^*)\}(1 - \alpha_j(\bar{\theta}_j))$ for all $\theta_i \leq -\theta_j^*$, $i = 1, 2$ (rather than $\alpha_i(\theta_i) = 0$ for $\theta_i \leq -\theta_j^*$). The modified mechanism satisfies EF and (ex post) IR as in the proof of Theorem 3. In addition, $t_i(\theta_i, \theta_j)$ is unchanged for all $\theta_i > -\theta_j^*$, and reporting $\hat{\theta}_i \leq -\theta_j^*$ continues to give payoff 0 in the worst case, so MMIC also follows as in the proof of Theorem 3.

Next, as $\alpha_i(\theta_i) > 0$ for all $\theta_i > \underline{\theta}_i$ in the modified mechanism, truthtelling is not weakly dominated for any type. In particular, truthtelling was weakly dominated only for types $\theta_i \leq -\theta_j^*$ in the unmodified $\alpha_i(\theta_i)$ double auction, and in the modified mechanism such a type does strictly better from truthtelling than from misreporting as type $\hat{\theta}_i < \theta_i$ against any opposing type $\theta_j \in (-\theta_i, -\hat{\theta}_i)$ (and $t_i(\theta_i, \theta_j)$ is unchanged for all $\theta_i > -\theta_j^*$, so truthtelling does not become weakly dominated for any of these types).

Finally, the argument for WBB is also similar to the proof of Theorem 3. More specifically, as in that proof, consider three cases:

If $\theta_1 < -\underline{\theta}_2$ and $\theta_2 < -\underline{\theta}_1$, then as in the proof of Theorem 3, WBB holds if and only if $\alpha_1(\theta_1) + \alpha_2(\theta_2) \leq 1$. If $\theta_i > -\theta_j^*$ for $i = 1, 2$, this holds because $\alpha_1(\bar{\theta}_1) + \alpha_2(\bar{\theta}_2) \leq 1$ (recalling that $\alpha_i(\theta_i)$ is nondecreasing in the unmodified $\alpha_i(\theta_i)$ double auction). If $\theta_1 > -\theta_2^*$ and $\theta_2 \leq -\theta_1^*$, it holds because

$$\alpha_1(\theta_1) + \alpha_2(\theta_2) \leq \tfrac{1}{2}(1 - \alpha_2(\bar{\theta}_2)) + \alpha_2(\theta_2) \leq \tfrac{1}{2} + \tfrac{1}{2}\alpha_2(\bar{\theta}_2) < 1,$$

and if $\theta_i \leq -\theta_j^*$ for $i = 1, 2$, it holds because

$$\alpha_1(\theta_1) + \alpha_2(\theta_2) \leq \tfrac{1}{2}(1 - \alpha_2(\bar{\theta}_2)) + \tfrac{1}{2}(1 - \alpha_1(\bar{\theta}_1)) < 1.$$

If $\theta_1 \geq -\underline{\theta}_2$ and $\theta_2 > -\underline{\theta}_1$, then a fortiori $\theta_1 > -\theta_2^*$ and $\theta_2 > -\theta_1^*$, so the argument is exactly as in the proof of Theorem 3.

Last, if $\theta_1 < -\underline{\theta}_2$ and $\theta_2 \geq -\underline{\theta}_1$, then, as in the proof of Theorem 3, WBB reduces to (10). If $\theta_1 > -\theta_2^*$, then the argument is exactly as in the proof of Theorem 3. If instead

$\theta_1 \le -\theta_2^*$, then (10) becomes

$$\left[\min\left\{\frac{1}{2}, \frac{\theta_1 - \underline{\theta}_1}{\bar{\theta}_2 - \theta_2^*}\right\} - \frac{\theta_1 - \underline{\theta}_1}{\bar{\theta}_2 + \theta_1}\right](1 - \alpha_2(\bar{\theta}_2)) \le 0,$$

which holds as the term in brackets is nonpositive.

Proof of Proposition 3

It is clear that the mechanism satisfies EF, (ex post) IR, and SBB. I now verify MMIC for an arbitrary buyer type v. (The argument for the seller is symmetric.)

If $\hat{v} \le c^*$, then $U_b(\hat{v}, c^*; v) = 0$, so IR and $\delta_{c^*} \in \Phi_s$ yield $U_b(v) \ge 0 \ge \inf_{\phi_s \in \Phi_s} U_b(\hat{v}, \phi_s; v)$.

If $\hat{v} > v$, then the observation that $t_b(v, c)$ is nonincreasing in v in the region where $v \ge c$ implies that $U_b(v, c; v) \ge U_b(\hat{v}, c; v)$ for all c, by the same argument as in the proof of Theorem 3 (Claim 3, Case (i)). Hence, $U_b(v) \ge \inf_{\phi_s \in \Phi_s} U_b(\hat{v}, \phi_s; v)$. This completes the proof for MMIC for $v \le c^*$, so assume henceforth that $v > c^*$.

If $\hat{v} \in (c^*, v)$, then $\hat{v} > p^*$. Hence, $U_b(\hat{v}, c; v) \le \tilde{U}_b(\hat{v}, c; v)$ for all c, where $\tilde{U}_b$ denotes utility under a standard reference rule (with $\varepsilon = 0$). Recalling that $\tilde{U}_b(v) \ge \inf_{\phi_s \in \Phi_s} \tilde{U}_b(\hat{v}, \phi_s; v)$ by Proposition 1, I complete the proof of MMIC by showing that $U_b(v) = \tilde{U}_b(v)$.

I claim that

$$\inf_{\phi_s \in \Phi_s} U_b(\hat{v}, \phi_s; v) = \frac{v - c^*}{v}(v - p^*)$$

for all $v > c^*$, whenever $\varepsilon < p^*$. In particular, I show that the infimum of $U_b(\hat{v}, \phi_s; v)$ over all ϕ_s with expectation c^* (and hence over Φ_s) is attained at $\phi_s = \delta_{0,v}^s$ (which is indeed an element of Φ_s).

To see this, first note that any ϕ_s with expectation c^* must put positive mass on the interval $[0, v)$. Now a type v buyer gets positive payoff against all seller types $c < v$ and gets payoff zero against all types $c \ge v$, so if ϕ_s puts positive mass on types $c > v$, then there exists another distribution ϕ_s' with the same mean that shifts this mass to $-\theta_i$ and reduces the mass on $[0, v)$, thus reducing buyer type v's payoff. Next, if ϕ_s puts positive mass on $(0, p^*]$, then shifting this mass to $c = 0$ and correspondingly increasing the mass on $(p^*, 1]$ weakly decreases the probability of trade and strictly decreases θ_i's expected transfer when trade occurs, so this modification also strictly reduces buyer type v's payoff.

Finally, if ϕ_s puts positive mass on (p^*, v), with $E^{\phi_s}[c|\theta \in (p^*, v)] = c$, then it is worse for buyer type v to split this mass between 0 and v in the proportions that preserve the mean. To see this, note that buyer type v's payoff against $c \in (p^*, v)$ is $(1 - \varepsilon)(v - c)$, while the type's payoff from facing a cost 0 seller with probability $(v - c)/v$ is $((v - c)/v)(v - p^*)$. Now $((v - c)/v)(v - p^*)$ is less than $(1 - \varepsilon)(v - c)$ if and only $\varepsilon < (1 - v + p^*)/v$, and a sufficient condition for this inequality to hold is $\varepsilon < p^*$. Thus, any measure ϕ_s that puts positive mass on the intervals $(0, p^*]$, (p^*, v), or $(v, 1]$ can be improved upon, so it follows that the infimum of $U_b(\hat{v}, \phi_s; v)$ over measures ϕ_s with expectation c^* is attained at the measure that puts mass only on $\{0, v\}$, namely $\delta_{0,v}$.

The above claim gives $U_b(v) = ((v - c^*)/v)(v - p^*)$. Since a standard reference rule corresponds to $\varepsilon = 0$, the same argument gives $\tilde{U}_b(v) = ((v - c^*)/v)(v - p^*)$. Hence, $U_b(v) = \tilde{U}_b(v)$, completing the proof of MMIC.

Finally, to see that truthtelling is not weakly dominated for any type when $\varepsilon > 0$, note that reporting $\hat{v} < v$ cannot dominate truthtelling because $U_b(v, (v + \hat{v})/2; v) > 0 = U_b(\hat{v}, (v + \hat{v})/2; v)$. Moreover, reporting $\hat{v} > v$ cannot dominate truthtelling because in this case $U_b(\hat{v}, c; v) \leq U_b(v, c; v)$ for all c, as noted above.

References

Agranov, Marina and Pietro Ortoleva (forthcoming), "Stochastic choice and preferences for randomization." *Journal of Political Economy*. [976]

Ahn, David S. (2007), "Hierarchies of ambiguous beliefs." *Journal of Economic Theory*, 136, 286–301. [976]

Arrow, Kenneth (1979), "The property rights doctrine and demand revelation under incomplete information." In *Economics and Human Welfare* (Michael J. Boskin, ed.), 23–38, New York, Academic Press. [981, 986]

Bergemann, Dirk and Stephen Morris (2005), "Robust mechanism design." *Econometrica*, 73, 1771–1813. [992]

Bergemann, Dirk and Karl Schlag (2011), "Robust monopoly pricing." *Journal of Economic Theory*, 146, 2527–2543. [992]

Bergemann, Dirk and Karl H. Schlag (2008), "Pricing without priors." *Journal of the European Economic Association*, 6, 560–569. [992]

Bewley, Truman F. (2002), "Knightian decision theory: Part I." Unpublished paper. [975]

Bodoh-Creed, Aaron (2014), "Correction to ambiguous beliefs and mechanism design." Working paper. [979]

Bodoh-Creed, Aaron L. (2012), "Ambiguous beliefs and mechanism design." *Games and Economic Behavior*, 75, 518–537. [975, 976, 977, 979, 980, 981]

Börgers, Tilman and Doug Smith (2012), "Robustly ranking mechanisms." *American Economic Review*, 102, 325–329. [987]

Bose, Subir and Arup Daripa (2009), "A dynamic mechanism and surplus extraction under ambiguity." *Journal of Economic Theory*, 144, 2084–2114. [974]

Bose, Subir and Suresh Mutuswami (2012), "Bilateral bargaining in an ambiguous environment." University of Leicester Department of Economics Working Paper 12/10. [974]

Bose, Subir, Emre Ozdenoren, and Andreas Pape (2006), "Optimal auctions with ambiguity." *Theoretical Economics*, 1, 411–438. [975, 976, 977, 979]

Bose, Subir and Ludovic Renou (2014), "Mechanism design with ambiguous communication devices." *Econometrica*, 82, 1853–1872. [974]

Carbajal, Juan Carlos and Jeffrey C. Ely (2013), "Mechanism design without revenue equivalence." *Journal of Economic Theory*, 148, 104–133. [979]

Cerreia-Vioglio, Simone, David Dillenberger, Pietro Ortoleva, and Gil Riella (2015), "Deliberately stochastic." Unpublished paper. [976]

Chatterjee, Kalyan and William Samuelson (1983), "Bargaining under incomplete information." *Operations Research*, 31, 835–851. [975]

Chung, Kim-Sau and Jeff Ely (2007), "Foundations of dominant strategy mechanisms." *Review of Economic Studies*, 74, 447–476. [992]

Crémer, Jacques and Richard P. McLean (1985), "Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent." *Econometrica*, 53, 345–361. [972]

Crémer, Jacques and Richard P. McLean (1988), "Full extraction of the surplus in Bayesian and dominant strategy auctions." *Econometrica*, 56, 1247–1257. [972]

d'Aspremont, Claude and Louis-André Gérard-Varet (1979), "Incentives and incomplete information." *Journal of Public Economics*, 11, 25–45. [981, 986]

De Castro, Luciano and Nicholas Yannelis (2010), "Ambiguity aversion solves the conflict between efficiency and incentive compatibility." Discussion Paper 1532, Center for Mathematical Studies in Economics and Management Science. [974, 976, 982, 983]

Di Tillio, Alfredo, Nenad Kos, and Matthias Messner (2014), "The design of ambiguous mechanisms." Working paper, Bocconi University. [974, 976]

Erdil, Aytek and Paul Klemperer (2010), "A new payment rule for core-selecting package auctions." *Journal of the European Economic Association*, 8, 537–547. [987]

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki (2015), "Stochastic choice and revealed perturbed utility." [976]

Garratt, Rod and Mark Pycia (2015), "Efficient bilateral trade." Working paper, available at SSRN 2444279. [975]

Gilboa, Itzhak (2009), *Theory of Decision Under Uncertainty*. Econometric Society Monographs. Cambridge University Press, New York. [972]

Gilboa, Itzhak and David Schmeidler (1989), "Maxmin expected utility with a non-unique prior." *Journal of Mathematical Economics*, 18, 141–153. [971, 976]

Grimmett, Geoffrey and David Stirzaker (2001), *Probability and Random Processes*, third edition. Oxford University Press, New York. [993]

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian persuasion." *American Economic Review*, 101, 2590–2615. [989]

Kos, Nenand and Matthias Messner (2013), “Extremal incentive compatible transfers.” *Journal of Economic Theory*, 148, 134–164. [979]

Krishna, Vijay and Motty Perry (2000), “Efficient mechanism design.” Working paper, available at SSRN 64934. [981, 986]

Linhart, P. B. and Roy Radner (1989), “Minimax-regret strategies for bargaining over several variables.” *Journal of Economic Theory*, 48, 152–178. [992]

Lopomo, Giuseppe, Luca Rigotti, and Chris Shannon (2014), “Uncertainty in mechanism design.” Unpublished paper, University of Pittsburgh. [975]

Machina, Mark J. (1985), “Stochastic choice functions generated from deterministic preferences over lotteries.” *Economic Journal*, 95, 575–594. [976]

Mailath, George and Andrew Postlewaite (1990), “Asymmetric information bargaining problems with many agents.” *Review of Economic Studies*, 57, 351–367. [991]

Makowski, Louis and Claudio Mezzetti (1994), “Bayesian and weakly robust first best mechanisms: Characterizations.” *Journal of Economic Theory*, 64, 500–519. [981, 986]

McAfee, R. Preston and Philip J. Reny (1992), “Correlated information and mechanism design.” *Econometrica*, 60, 395–421. [972]

Milgrom, Paul R. and Ilya Segal (2002), “Envelope theorems for arbitrary choice sets.” *Econometrica*, 70, 583–601. [975, 979, 980]

Myerson, Roger B. (1981), “Optimal auction design.” *Mathematics of Operations Research*, 6, 58–73. [977]

Myerson, Roger B. and Mark A. Satterthwaite (1983), “Efficient mechanisms for bilateral trading.” *Journal of Economic Theory*, 29, 265–281. [971, 972]

Raiffa, Howard (1961), “Risk, ambiguity, and the Savage axioms: Comment.” *Quarterly Journal of Economics*, 75, 690–694. [976]

Saito, Kota (2015), “Preferences for flexibility and randomization under uncertainty.” *American Economic Review*, 105, 1246–1271. [976]

Schmeidler, David (1989), “Subjective probability and expected utility without additivity.” *Econometrica*, 57, 571–587. [976]

Segal, Ilya and Michael D. Whinston (2002), “The mirrlees approach to mechanism design with renegotiation (with applications to hold-up and risk-sharing).” *Econometrica*, 70, 1–45. [979]

Shmaya, Eran and Leeat Yariv (2009), “Foundations for Bayesian updating.” Unpublished paper, CalTech. [989]

Williams, Steven R. (1999), “A characterization of efficient, Bayesian incentive compatible mechanisms.” *Economic Theory*, 14, 155–180. [981, 986]

Wilson, Robert (1989), “Game theoretic analysis of trading processes.” In *Advances in Economic Theory* (Truman Bewley, ed.), 33–70, Cambridge University Press, New York. [971]

Wolitzky, Alexander (2016), “Mechanism design with maxmin agents: Theory and an application to bilateral trade.” Unpublished paper. [976]

Co-editor Faruk Gul handled this manuscript.

Submitted 2015-1-29. Final version accepted 2015-10-15. Available online 2015-10-15.