American Economic Journal: Microeconomics 2021, 13(1): 148–178
https://doi.org/10.1257/mic.20180184

Full Implementation under Ambiguity[†]

By Huiyi Guo and Nicholas C. Yannelis*

This paper introduces the maxmin expected utility framework into the problem of fully implementing a social choice set as ambiguous equilibria. Our model incorporates the Bayesian framework and the Wald-type maxmin preferences as special cases and provides insights beyond the Bayesian implementation literature. We establish necessary and almost sufficient conditions for a social choice set to be fully implementable. Under the Wald-type maxmin preferences, we provide easy-to-check sufficient conditions for implementation. As applications, we implement the set of ambiguous Pareto-efficient and individually rational social choice functions, the maxmin core, the maxmin weak core, and the maxmin value. (*JEL* D71, D81, D82)

In implementation theory, a mechanism designer aims to elicit information from agents and realize an exogenous social choice set or function. If a mechanism can be designed such that all its equilibria coincide with the social choice set, then the set is said to be fully implementable. When agents have private information, the subjective expected utility framework has been widely adopted in the literature to model agents' preferences. However, we have known from Ellsberg (1961) that the subjective expected utility hypothesis is problematic. To this end, nonexpected utility decision theory has been developed.[1] In particular, the seminal work of Gilboa and Schmeidler (1989) proposes the maxmin expected utility, which is one of the successful alternatives in describing agents' decision-making under ambiguity. When maxmin expected utility is adopted, new insights have emerged in different mechanism design problems. However, the full implementation problem has not been considered yet under maxmin preferences.

By assuming that agents are maxmin expected utility maximizers, we provide a new framework to study full implementation. Gilboa and Schmeidler's (1989) maxmin expected utility postulates that a decision-maker may have multiple beliefs about the underlying state of the world and makes decisions with the worst-case belief. In our asymmetric information environment, we assume that agents know

*Guo: Department of Economics, Texas A&M University, 4228 TAMU, College Station, TX 77843 (email: huiyiguo@tamu.edu); Yannelis: Department of Economics, The University of Iowa, W380 John Pappajohn Business Building, Iowa City, IA 52242 (email: nicholasyannelis@gmail.com). Leslie Marx was coeditor for this article. We thank a referee whose comments helped us improve the paper, as well as many conference participants. An earlier version of this paper is a chapter of Huiyi Guo's doctoral thesis at the University of Iowa. Huiyi Guo acknowledges the financial support of Ballard and Seashore Dissertation Fellowship from the University of Iowa.

[†]Go to https://doi.org/10.1257/mic.20180184 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] See, e.g., Wald (1945); Bewley (2002); Gilboa and Schmeidler (1989); Maccheroni, Marinacci, and Rustichini (2006); Cerreia-Vioglio et al. (2011).

little about each other's private information and thus may form ambiguous beliefs about the private information held by others. As special cases, this setup includes both the Bayesian framework, where each multi-belief set is a singleton, and the Wald-type maxmin preferences, where each agent's decision-making is based on the worst-case information of other agents.

The solution concept we adopt is essentially the Bayesian (Nash) equilibrium with ambiguous beliefs, which we call an ambiguous equilibrium. Echoing the results in the Bayesian implementation literature, we show that the conditions of ambiguous incentive compatibility and ambiguous monotonicity are necessary and almost sufficient for a social choice set to be fully implementable as ambiguous equilibria. When establishing sufficient conditions for full implementation, we strengthen the two key conditions by imposing a bad outcome property. The bad outcome property is usually a weak requirement in environments like exchange economies or transferable utility environments.

The mechanism we construct to implement a social choice set differs from the ones in the Bayesian implementation literature. In the Bayesian framework, each agent's belief satisfies a full support assumption. This fact has been utilized in the proof to achieve full implementation of social choice sets in the literature. However, with ambiguous beliefs, especially those under the Wald-type maxmin preferences, there may exist non-full-support beliefs in the ambiguous belief set. This necessitates a different construction to establish full implementation under ambiguity. Although the seminal works by Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989a), and Jackson (1991) have been useful in carrying out the new construction, the details and arguments are nontrivial.

In exchange economies, when agents have Wald-type maxmin preferences and private value utility functions, we provide easy-to-check and also weak conditions that are sufficient for full implementation. This contrasts with the fact that the conditions for Bayesian implementation are usually either relatively demanding or difficult to check. De Castro and Yannelis (2018) have shown that each ambiguous Pareto-efficient social choice function is ambiguous incentive compatible. We further find that if the social choice set is ambiguous Pareto efficient and every unacceptable deception profile lowers at least one agent's interim utility, then the social choice set also satisfies the ambiguous monotonicity condition. At last, the ambiguous individual rationality condition is sufficient to guarantee the bad outcome property when agents have nonzero initial endowments.

By applying the simplified sufficient conditions, we are able to implement a few solution concepts in exchange economies that are not only of interest but also may not be implemented under a Bayesian framework. In particular, under the Wald-type maxmin preferences and private value utility functions, we show that the set of all ambiguous Pareto-efficient and individually rational social choice functions is fully implementable as ambiguous equilibria. This extends the result of de Castro, Liu, and Yannelis (2017a, b) on partial implementation of a social choice function to full implementation of a social choice set. Also, we show that the maxmin core of de Castro, Pesce, and Yannelis (2011), the maxmin weak core, and the maxmin value of Angelopoulos and Koutsougeras (2015) are fully implementable as ambiguous equilibria. This contrasts with the Bayesian framework, under which notions like

efficient social choice sets and the core are generally not implementable (see, for example, Palfrey and Srivastava 1987).

Our paper is related to two strands of the literature. The first one is on mechanism design with ambiguity-averse agents. Instead of fully implementing a social choice set, these papers implement a given social choice function and impose some assumptions on equilibrium selection. Under the Wald-type maxmin preferences, de Castro and Yannelis (2018) prove that every Pareto-efficient social choice function is incentive compatible. De Castro, Liu, and Yannelis (2017a, b) and Liu (2016) partially implement efficient social choice functions as maxmin equilibria. Their maxmin equilibrium is different from the ambiguous equilibrium in the current paper: their agents make decisions based on opponents' worst-case information and worst-case strategies, but we assume that agents anticipate opponents to use equilibrium strategies. A few other partial implementation papers, for example, Bose and Renou (2014), Wolitzky (2016), and Guo (2019), adopt solution concepts that are essentially the ambiguous equilibrium and allow for maxmin preferences that are not Wald type. In these various setups, ambiguity aversion can help the designer soften the conflict between efficiency and incentive compatibility, although the conflict may still exist. There are also papers studying revenue maximization with ambiguity-averse agents, e.g., Bodoh-Creed (2012) and di Tillio, Kos, and Messner (2017). All the abovementioned papers focus on incentive compatibility, and none is concerned with the issue of multiple equilibria. Nonetheless, the current paper studies fully implementing an exogenous social choice set as ambiguous equilibria and thus is different from the abovementioned works. To the best of our knowledge, no similar results to ours exist in the literature.

The second strand of the literature is full implementation. The problem of full implementation has been studied extensively in both a complete information environment and in one with asymmetric information. With complete information, Maskin (1999), Saijo (1988), Repullo (1987), and Dutta and Sen (1991) among others show that a monotonicity condition is the key condition for Nash implementation. With asymmetric information, the Bayesian implementation literature (e.g., Postlewaite and Schmeidler 1986; Palfrey and Srivastava 1987, 1989a, b; and Jackson 1991) has established necessary and almost sufficient conditions to implement a social choice set as Bayesian equilibria. The current paper differs from the abovementioned works in the maxmin expected utility setup, in implications, and in the construction of the mechanism. From the perspective of setup, this paper provides a unified treatment for full implementation with maxmin preferences. In particular, the benchmark models of the Bayesian framework and the Wald-type maxmin framework are covered as special cases. In terms of implications, under the Wald-type maxmin preferences, we provide weak and easy-to-check sufficient conditions for full implementation. This contrasts with the Bayesian framework under which the conditions for implementation are often demanding or difficult to check. Solution concepts including the ambiguous Pareto-efficient and individually rational social choice set, the maxmin core, the maxmin weak core, and the maxmin value are implementable under the Wald-type maxmin framework, while various efficient social choice sets have been demonstrated by Palfrey and Srivastava (1987) to be nonimplementable in a Bayesian framework. From the viewpoint of technical

details in the construction of our mechanism, we do not assume that every belief in the ambiguous belief set has full support, which requires the design of new mechanisms for full implementation under ambiguity.

The paper proceeds as follows. Section I presents the primitives of the paper and introduces key conditions for full implementation. In Section II, we establish the necessity and almost sufficiency of the key conditions. Section III adopts the Wald-type maxmin preferences and provides easy-to-check conditions for full implementation. Several applications in exchange economies are provided in this section. Section IV concludes. All proofs are relegated to the Appendix.

I. The Model

A. *Environment*

Consider an environment with a finite **set of agents** $I = \{1, \ldots, n\}$.

Each agent i's private information is summarized by a **type** $t_i \in T_i$. We focus on the case that each type set T_i is finite to avoid technical complication. The set of all type profiles is denoted by $T = \prod_{i \in I} T_i$, where a generic element is $t = (t_i)_{i \in I}$. Similarly, for an agent i, denote the set of all others' type profiles by $T_{-i} = \prod_{j \neq i} T_j$, where a generic element is $t_{-i} = (t_j)_{j \neq i}$.

Following Gilboa and Schmeidler (1989), we assume that agents have multiple probability assessments toward others' types. Each agent i with type t_i has an **ambiguous belief** $\Pi_i(t_i)$, where the function $\Pi_i : T_i \to 2^{\Delta(T_{-i})}$ maps each type of agent i into a nonempty, compact, and convex set of distributions over T_{-i}. An element $\pi_i(t_i) \in \Pi_i(t_i)$ assigns probability $\pi_i(t_i)[t_{-i}]$ to the event that others have type profile t_{-i}.

When each $\Pi_i(t_i)$ is a singleton, the system of ambiguous beliefs degenerates to the Bayesian case as in Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989a), Jackson (1991), etc., where every agent's interim belief is updated from a prior of himself or a prior shared among all agents.

Let A denote **the set of feasible outcomes**, i.e., the set of all lotteries over a pure feasible outcome set X. A **social choice function** is a mapping $f : T \to A$. A **social choice set** is a set of social choice functions.

Agent i's **utility function** $u_i : X \times T \to \mathbb{R}$ defines his utility of receiving a pure outcome $a \in X$ when the realized type profile is $t \in T$. To accommodate lotteries whose realizations follow objective distributions, we extend the domain of u_i to $A \times T$ and define the payoff from a lottery using the expected utility.

As in Gilboa and Schmeidler (1989), each type-t_i agent i's interim payoff from a social choice function f takes the form of the **maxmin expected utility**,

$$\min_{\pi_i(t_i) \in \Pi_i(t_i)} \sum_{t_{-i} \in T_{-i}} u_i\big(f(t), t\big)\, \pi_i(t_i)[t_{-i}].$$

Namely, an agent takes into account the worst-case belief toward other agents' private information when evaluating his interim payoff. When ambiguous beliefs

are singletons, the interim preferences are consistent with the subjective expected utility theory adopted to study Bayesian implementation.

Subsequently, we introduce two assumptions on ambiguous beliefs. We will be explicit when imposing either of these assumptions. Each assumption below is stated for type $t_i \in T_i$ of agent $i \in I$. When T_{-i} is not a singleton, Assumptions 1 and 2 are incompatible.

The first assumption assumes that every distribution in the ambiguous belief of type t_i has full support.

ASSUMPTION 1: $\pi_i(t_i)[t_{-i}] > 0$ *for all* $t_{-i} \in T_{-i}$ *and* $\pi_i(t_i) \in \Pi_i(t_i)$.

The second assumption postulates that type-t_i agent i is extremely ambiguity averse: he believes that all distributions over the set T_{-i} are possible.

ASSUMPTION 2: $\Pi_i(t_i) = \Delta(T_{-i})$.

When Assumption 2 holds for type t_i of agent i, we say this type exhibits **Wald-type maxmin preference**. This type's maxmin expected utility takes a particularly simple form:

$$\min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\big(f(t),t\big)\,\pi_i(t_i)[t_{-i}] = \min_{t_{-i}\in T_{-i}} u_i\big(f(t),t\big).$$

Namely, type-t_i agent i makes decisions based on the worst-case type profile of other agents.

B. *Implementation*

A mechanism designer aims to implement an exogenously given social choice set F. A **mechanism** is a pair $(M,g) = \big(\prod_{i\in I} M_i, g\big)$, where M_i is the set of all messages that agent i can submit to the mechanism designer. We call $M = \prod_{i\in I} M_i$ the **message space**. When $M = T$, the mechanism is a direct mechanism. However, we follow the full implementation literature and adopt a general message space to achieve full implementation more easily. An **outcome function** is a mapping $g : M \to A$, which assigns a feasible outcome to each message profile. Agent i's **strategy** $\sigma_i : T_i \to M_i$ is a private information contingent plan of submitting messages. A **strategy profile** of all agents is given by $\sigma = (\sigma_i)_{i\in I}$. Similarly, let σ_{-i} denote the strategy profile of all agents except i. Following most of the papers on Bayesian implementation and on mechanism design with ambiguity-averse agents, we restrict attention to pure strategies.[2]

The solution concept we adopt in the paper is an adaptation of the Bayesian equilibrium with ambiguous beliefs. We call it ambiguous equilibrium to differentiate from the Bayesian equilibrium.

[2] See, for example, Postlewaite and Schmeidler (1986); Palfrey and Srivastava (1989a); Jackson (1991); di Tillio, Kos, and Messner (2017); Wolitzky (2016); and de Castro and Yannelis (2018).

DEFINITION 1: *A strategy profile* σ^* *is an* ***ambiguous equilibrium*** *of the mechanism* (M, g) *if*

$$\min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\big(g\big(\sigma^*(t)\big), t\big)\, \pi_i(t_i)\big[t_{-i}\big] \geq \min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\big(g\big(\sigma_i'(t_i), \sigma_{-i}^*(t_{-i})\big), t\big)\, \pi_i(t_i)\big[t_{-i}\big]$$

for all $i \in I$, $t_i \in T_i$, *and* $\sigma_i' : T_i \to M_i$.

A mechanism (M, g) **fully implements** a social choice set F as ambiguous equilibria if the following two conditions are satisfied:

(i) for any $f \in F$, there exists an ambiguous equilibrium σ of the mechanism (M, g) such that $g\big(\sigma(t)\big) = f(t)$ for all $t \in T$;

(ii) if σ is an ambiguous equilibrium of the mechanism (M, g), then there exists $f \in F$ such that $g\big(\sigma(t)\big) = f(t)$ for all $t \in T$.

If the first requirement is satisfied, then the social choice set F is said to be **partially implemented** by (M, g). In this case, for each function in the social choice set, there exists a "good" equilibrium leading to consistent outcomes. If the second requirement is satisfied, then there does not exist any "bad" equilibrium, which leads to outcomes inconsistent with functions in the social choice set.

C. *Conditions*

In this subsection, we introduce conditions that are useful for full implementation under ambiguity. They are the ambiguous incentive compatibility condition, the ambiguous monotonicity condition, and the bad outcome property. These conditions are defined for social choice sets. When a singleton social choice set $F = \{f\}$ satisfies these conditions, we say that the social choice function f satisfies these conditions.

Our first condition is a version of the incentive compatibility condition. It says that for each $f \in F$, truthful reporting is an ambiguous equilibrium in the direct mechanism with outcome function f. Formally, the definition is presented as follows.

DEFINITION 2: *A social choice set* F *is said to satisfy the* ***ambiguous incentive compatibility*** *condition if*

$$\min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\big(f(t), t\big)\, \pi_i(t_i)\big[t_{-i}\big] \geq \min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\big(f(t_i', t_{-i}), t\big)\, \pi_i(t_i)\big[t_{-i}\big]$$

for all $f \in F$, $i \in I$, *and* $t_i, t_i' \in T_i$.

This condition guarantees the existence of good equilibria, i.e., equilibria that are consistent with functions in the social choice set.

To introduce the ambiguous monotonicity condition, we first define the notion of deceptions, which can be viewed as strategies in direct mechanisms. We remark that the mechanism we design for implementation is not a direct mechanism, but the notion of deception plays an important role in the analysis. A **deception** of agent i is a mapping $\alpha_i : T_i \rightarrow T_i$. Under α_i, type-t_i agent i reports $\alpha_i(t_i)$ to the mechanism designer. Specifically, when $\alpha_i : T_i \rightarrow T_i$ is the identity mapping, it represents truthful reporting of private information. We denote by α the deception profile $(\alpha_i)_{i \in I}$ and by α_{-i} the profile $(\alpha_j)_{j \neq i}$.

Given a social choice set F and a social choice function $f \in F$, the deception profile α is **acceptable** if there exists $f' \in F$ such that $f'(t) = f(\alpha(t))$ for all $t \in T$. Otherwise, the deception profile is **unacceptable**.

For a social choice function $f \in F$, an agent i, and a type $t_i' \in T_i$, let the set $H^f_{t_i'}$ be the collection of all social choice functions $h : T \rightarrow A$ such that

$$\min_{\pi_i(t_i) \in \Pi_i(t_i)} \sum_{t_{-i} \in T_{-i}} u_i\big(f(t), t\big)\, \pi_i(t_i)\big[t_{-i}\big] \geq \min_{\pi_i(t_i) \in \Pi_i(t_i)} \sum_{t_{-i} \in T_{-i}} u_i\big(h(t_i', t_{-i}), t\big)\, \pi_i(t_i)\big[t_{-i}\big], \quad \forall t_i \in T_i.$$

The set $H^f_{t_i'}$ is called a **reward set**, and a function in it is called a **reward function**.

The ambiguous monotonicity condition is presented below.

DEFINITION 3: *A social choice set F is said to satisfy the **ambiguous monotonicity** condition if for any social choice function $f \in F$ and unacceptable deception profile α, there exists $i \in I$, $t_i \in T_i$, and $h \in H^f_{\alpha_i(t_i)}$ such that*

$$\min_{\pi_i(t_i) \in \Pi_i(t_i)} \sum_{t_{-i} \in T_{-i}} u_i\big(h(\alpha(t)), t\big)\, \pi_i(t_i)\big[t_{-i}\big] > \min_{\pi_i(t_i) \in \Pi_i(t_i)} \sum_{t_{-i} \in T_{-i}} u_i\big(f(\alpha(t)), t\big)\, \pi_i(t_i)\big[t_{-i}\big].$$

The condition implies that whenever agents unanimously choose a social choice function and adopt an unacceptable deception profile, there is an agent who can benefit from deviating unilaterally and proposing a reward function. Such an agent is often called a "whistle-blower" since this agent can signal to the mechanism designer that a bad equilibrium is reached.

Along with the conditions of ambiguous incentive compatibility and ambiguous monotonicity, the following sufficient condition is imposed to fully implement a social choice set.

DEFINITION 4: *A social choice set F is said to satisfy the **bad outcome property** if there exists $\underline{a} \in A$ and $\delta > 0$ such that*

$$\min_{t_{-i} \in T_{-i}, t' \in T} u_i\big(f(t'), t\big) - \max_{t_{-i} \in T_{-i}} u_i(\underline{a}, t) \geq \delta$$

for all $f \in F$, $i \in I$, and $t_i \in T_i$.

With this property, there exists a bad outcome $\underline{a}$, whose maximum payoff to type-t_i agent i is lower than and bounded away from the worst-case payoff from any social choice outcome.

Our bad outcome property holds naturally for a few common social choice sets. In Section III, we will formally define an exchange economy where every agent has nonzero initial endowment and private evaluation. If we wish to implement an individually rational social choice set, then appropriating all initial endowments from all agents may serve as a bad outcome. In environments other than exchange economies, there may exist other bad outcomes. For instance, in a quasi-linear environment, appropriating a sufficiently large amount of money from all agents can usually serve as a bad outcome for any individually rational social choice set. In a matching environment, if all types of agents prefer being matched to other agents to being unmatched, then the outcome of no matching may serve as a bad outcome for Pareto-efficient social choice sets. In an election environment where all voters prefer having an elected candidate to having no leader, the outcome of having no leader could be a bad outcome for Pareto-efficient social choice sets.

II. Main Result

A. *Statement of the Theorem*

The following theorem is the main result of the paper. It provides conditions that are necessary and sufficient for fully implementing a social choice set under ambiguous beliefs in general environments.

THEOREM 1:

(*i*) *If a social choice set F is fully implementable as ambiguous equilibria, then F satisfies ambiguous incentive compatibility and ambiguous monotonicity.*

(*ii*) *For each $i \in I$, suppose every $t_i \in T_i$ satisfies either Assumption* 1 *or Assumption* 2*. If a social choice set F satisfies ambiguous incentive compatibility, ambiguous monotonicity, and the bad outcome property, then F is fully implementable as ambiguous equilibria.*

The formal proof of the theorem is relegated to the Appendix. We provide an idea of the proof in Section IIB.

Notice that the second part of the theorem does not rely on the cardinality restriction $n \geq 3$ commonly seen in the literature. This is because we adopt a stochastic mechanism for full implementation.

We assume that every type satisfies either Assumption 1 or Assumption 2 in the second part of the theorem. Thus, the closure condition in the implementation literature is automatically satisfied. Note that neither Assumption 1 nor Assumption 2 is needed for the first part of Theorem 1. Also, neither is needed for implementing a singleton social choice set. Thus, we also have the result below.

COROLLARY 1: *If a social choice function f satisfies ambiguous incentive compatibility, ambiguous monotonicity, and the bad outcome property, then f is fully implementable as ambiguous equilibria.*

B. *Discussion of the Proof*

The proof of the first part of the theorem is standard. The proof of the second part requires new arguments that are not found in the literature. We provide here a sketch of the proof of the second part.

We construct a mechanism (M, g) to implement F. Each agent i's message $m_i \in M_i$ has five components: the first component m_i^1 is a type of agent i; the second component m_i^2 proposes a social choice function in F; the third and fourth components m_i^3 and m_i^4 are nonnegative integers; the fifth component m_i^5 proposes a social choice function, which is not required to be in F. We partition the message space M into sets M^1, M^2, and M^3 as follows:

$$M^1 = \left\{m \in M \,|\, \exists f \in F \text{ such that } m_i^2 = f \text{ and } m_i^3 = 0,\ \forall i \in I\right\},$$

$$M^2 = \Big\{m \in M \,|\, \exists f \in F,\ i \in I,\ h \in H^f_{m_i^1} \text{ such that } m_i^3 > 0,\ m_i^5 = h,\ m_j^2 = f,\ m_j^3 = 0,\ \forall j \neq i\Big\},$$

$$M^3 = M \backslash \left\{M^1 \cup M^2\right\}.$$

Our mechanism incorporates a "bad lottery" that is formalized in the Appendix to dissolve bad equilibria. The lottery has two realizations: the bad outcome and an outcome that is better for every agent. The probability of the realization of the better outcome is increasing in $\sum_{j \in I} m_j^4$. If assigned the lottery, every i can benefit from reporting a larger m_i^4 to increase the probability of the realization of the better outcome. Thus, the bad lottery creates an "open set" over which agents have strict preferences so that best responses may not exist.

Our mechanism assigns the following rules. Rule 1 applies when the message profile $m \in M^1$. In this case, the reported types in the first components of the messages and the common social choice function in the second components determine the outcome. Rule 2 applies when $m \in M^2$. Let i denote the agent reporting $m_i^3 > 0$. The mechanism assigns a compound lottery between the reward function m_i^5 and the bad lottery constructed above. In the compound lottery, the probability of the realization of m_i^5 is increasing in m_i^3. Rule 3 is triggered in all other cases, where agents are assigned the bad lottery.

For each $f \in F$, Claim 1 in the Appendix establishes the existence of a good equilibrium that triggers Rule 1: every agent truthfully reports, proposes f, and submits zero in the first, second, and third components of his message. To see this, first notice that f satisfies ambiguous incentive compatibility, and thus no agent has the incentive to remain under Rule 1 and misrepresent his type. According to the definition of

reward functions and the bad lottery construction, no agent has the incentive to deviate and trigger Rule 2. Due to the bad lottery construction, no agent has the incentive to deviate and trigger Rule 3.

Claims 2 through 4 demonstrate that every ambiguous equilibrium in the mechanism is a good equilibrium.

Claim 2 first shows that every agent reports zero in the third component of the message in equilibrium. To see this, suppose that some type-t_i agent i reports $m_i^3 > 0$. Notice that for each t_{-i}, either Rule 2 or Rule 3 is triggered at t. In both cases, the bad lottery is used with positive probability. Thus, type-t_i agent i can benefit from deviating with a larger m_i^4 to increase the probability that the better outcome is realized in the bad lottery.

Claim 3 then shows that in any ambiguous equilibrium, agents agree on a common $f \in F$ in the second components of their messages. Otherwise, there exists $t^* \in T$ under which agents propose different social choice functions. One can fix any $i \in I$ and show that type-t_i^* agent i can profitably deviate by reporting a larger m_i^4. To see this, notice that for each $t_{-i} \in T_{-i}$, either Rule 1 or Rule 3 is triggered at (t_i^*, t_{-i}). We discuss the following two cases. First, when t_i^* satisfies Assumption 1, each distribution in the ambiguous belief has full support. As the deviation improves the outcome when Rule 3 is triggered and does not change the outcome when Rule 1 is triggered, the agent's interim payoff will increase. Second, when t_i^* satisfies Assumption 2, the agent's interim payoff is solely determined by the worst one or multiple $t_{-i} \in T_{-i}$. Due to the bad lottery design, the worst $t_{-i} \in T_{-i}$ triggers Rule 3. By deviating with a larger m_i^4, the bad lottery imposes more weight on the better outcome, which increases the worst-case payoff of t_i^*.

Claim 4 further shows that agents adopt an acceptable deception profile for a social choice function $f \in F$ in equilibrium, and thus F is fully implemented. Otherwise, some agent i can benefit from proposing a profitable reward function m_i^5 (by ambiguous monotonicity of F) and submitting a sufficiently large $m_i^3 > 0$. In this case, the outcome function approximates the reward function, and thus the deviation is profitable.

C. *Connection with the Literature*

Although our mechanism shares similarities with those in Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987), and Jackson (1991), their mechanisms do not work under the maxmin framework. This is because not all beliefs have full support in the maxmin framework. To prevent a bad strategy profile, i.e., one leading to outcomes that are inconsistent with F, from being an equilibrium, their mechanisms rely on the full support assumption. In particular, under some bad strategy profiles, there is a type-t_i agent i, who by deviating can weakly increase his ex post utility under all $t_{-i} \in T_{-i}$ and can strictly increase his utility under some $t_{-i} \in T_{-i}$. As a Bayesian agent with full-support belief, this type t_i agent i has the incentive to deviate, which prevents the bad strategy profile from being an equilibrium. However, when this agent has the Wald-type maxmin preference instead, his payoff is solely decided by the worst $t_{-i} \in T_{-i}$. Without being able to increase the ex post utility under all the worst $t_{-i} \in T_{-i}$, the agent does not have the incentive

to deviate. Thus, under the maxmin framework, there may exist bad equilibria in the mechanisms of Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987), and Jackson (1991).

Our stochastic mechanism is related to the one in Bergemann and Morris (2011), but their mechanism works for implementing social choice functions only. In particular, they do not need to argue how to dissolve the bad equilibria where agents propose to implement different social choice functions. However, this argument is needed to implement a social choice set and is challenging when agents do not have full-support beliefs. To make this argument go through, we introduce a stronger bad outcome property than theirs and assume that each type of each agent satisfies either Assumption 1 or Assumption 2. The case-by-case discussion in Claim 3 of the Appendix shows how we dissolve the bad equilibria where agents propose different social choice functions.

III. Wald-Type Maxmin Preferences: Applications

Throughout this section, we impose Assumption 2 on all types of all agents. Namely, we consider the Wald-type maxmin preferences. We also assume that agents have **private value** utility functions, i.e., $u_i\big(a,\big(t_i,t_{-i}\big)\big) = u_i\big(a,\big(t_i,t'_{-i}\big)\big)$ for all $a \in A$, $i \in I$, $t_i \in T_i$, and $t_{-i}, t'_{-i} \in T_{-i}$. Thus, we can denote the ex post utility for each type-t_i agent i to receive outcome a by $u_i\big(a,t_i\big)$ for simplicity. These assumptions are of interest because under them we can provide weak and easy-to-check sufficient conditions to guarantee ambiguous incentive compatibility and ambiguous monotonicity.

We also restrict our discussion to exchange economies in this section, mainly because the bad outcome property can be easily verified for several useful solution concepts defined in exchange economies in Sections IIIB through IIID. In an exchange economy, there are L goods, and the total amount of each good l is a positive number $e^l \in \mathbb{R}_{++}$. Each agent i is assumed to have a nonzero deterministic initial endowment: $e_i = \big(e_i^1, e_i^2, \ldots, e_i^L\big) \in \mathbb{R}_+^L \backslash \{\mathbf{0}\}$. Let $e = \big(e_i\big)_{i\in I}$ denote the no trade outcome. For each good, the aggregate initial endowment across agents is consistent with the total resource of this good, i.e., $\sum_{i\in I} e_i^l = e^l$ for each good $l = 1,\ldots,L$.

Agents cannot consume more than their aggregate initial endowment. Hence, the set of feasible pure outcomes, X, is defined by

$$X = \left\{ \big(x_i\big)_{i\in I} \,|\, x_i \in \mathbb{R}_+^L, \forall i \in I; \sum_{i\in I} x_i^l \leq \sum_{i\in I} e_i^l, \forall l = 1,2,\ldots,L \right\}.$$

The set of all feasible outcomes is $A = \Delta\big(X\big)$.

A coalition is a nonempty subset of I. When an outcome is feasible within a coalition, agents in the coalition cannot consume more than their aggregate initial endowment. Hence, for each coalition S, the set of pure outcomes that are feasible within S, X_S, is defined by

$$X_S = \left\{ \big(x_i\big)_{i\in I} \in X \Big| \sum_{i\in S} x_i^l \leq \sum_{i\in S} e_i^l, \forall l = 1,2,\ldots,L \right\}.$$

Define $A_S = \Delta(X_S)$ as the set of outcomes that are feasible within S. Notice that $X = X_I$ and $A = A_I$.

Since free disposal is allowed, the zero outcome $\mathbf{0} \in \mathbb{R}_+^{nL}$ is a feasible pure outcome. When defined on pure outcomes, each agent's utility function is assumed to be increasing and continuous in each dimension of his private consumption and independent of others' consumption.

We now introduce ambiguous Pareto efficiency, adopting the interim dominance notion of Holmström and Myerson (1983). Under this notion, a social choice function dominates the other if the first function is at least as good as the second one for all agents and all types, and at least one agent prefers the first function under one of his types.

DEFINITION 5: *A social choice set F is said to satisfy the **ambiguous Pareto efficiency** condition if there does not exist a social choice function $f \in F$ and another social choice function $y : T \to A$ such that*

$$\min_{t_{-i} \in T_{-i}} u_i\big(y(t), t_i\big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f(t), t_i\big)$$

for all $i \in I$ and $t_i \in T_i$, and the strict inequality holds for some $i \in I$ and $t_i \in T_i$.

Below, we define the ambiguous individual rationality condition.

DEFINITION 6: *A social choice set F is said to satisfy the **ambiguous individual rationality** condition if*

$$\min_{t_{-i} \in T_{-i}} u_i\big(f(t_i, t_{-i}), t_i\big) \geq u_i\big(e, t_i\big)$$

for all $f \in F$, $i \in I$, and $t_i \in T_i$.

In fact, it is easy to see that F satisfies the ambiguous individual rationality condition if and only if

$$u_i\big(f(t_i, t_{-i}), t_i\big) \geq u_i\big(e, t_i\big)$$

for all $f \in F, i \in I, t_i \in T_i$, and $t_{-i} \in T_{-i}$. This means that F is ambiguous individually rational if and only if it is ex post individually rational. In the rest of the paper, we choose to adopt the term of ambiguous individual rationality because all notions considered in Sections IIIB through IIID are interim notions.

A. *Easy-to-Check Sufficient Conditions*

De Castro and Yannelis (2018) have shown that every ambiguous Pareto-efficient social choice function is ambiguous incentive compatible if and only if agents have Wald-type maxmin preferences. This means that the conflict between efficiency and incentive compatibility is resolved under Wald-type maxmin preferences. Their result is presented below.

LEMMA 1 (de Castro and Yannelis 2018): *Every ambiguous Pareto-efficient social choice function f satisfies the ambiguous incentive compatibility condition.*

The intuition can be captured by the following two-by-one example. Suppose that agent 1's types are labeled by t_1^1 and t_1^2 and that agent 2's type is t_2. When type-t_1^1 agent 1 prefers to misreport t_1^2, one can define a new social choice function y by $y(t_1^1, t_2) = y(t_1^2, t_2) = f(t_1^2, t_2)$. The new social choice function y makes type-t_1^1 agent 1 better off compared to f in the interim stage and preserves his interim payoff under t_1^2. Also, type-t_2 agent 2 is not worse off by receiving y. To see this, the sure outcome of $y(t_1^1, t_2) = y(t_1^2, t_2) = f(t_1^2, t_2)$ is assigned under the new function y, and the uncertain outcome $f(t_1^1, t_2)$ or $f(t_1^2, t_2)$ is assigned under f. As agent 2 has the Wald-type maxmin preference and private valuation, y is at least as good as f to him. This contradicts the fact that f is ambiguous Pareto efficient. Hence, an ambiguous Pareto-efficient social choice function is ambiguous incentive compatible.

We also establish a weak condition that is sufficient for ambiguous monotonicity. If a social choice set is ambiguous Pareto efficient and every unacceptable deception profile lowers at least one agent's interim payoff, then the ambiguous monotonicity condition is satisfied.

LEMMA 2: *Let F be an ambiguous Pareto-efficient social choice set. If for any social choice function $f \in F$ and unacceptable deception profile α, there exists an agent $i \in I$ and type $t_i^* \in T_i$ such that*

$$\min_{t_{-i} \in T_{-i}} u_i\big(f\big(t_i^*, t_{-i}\big), t_i^*\big) > \min_{t_{-i} \in T_{-i}} u_i\big(f\big(\alpha\big(t_i^*, t_{-i}\big)\big), t_i^*\big),$$

then the social choice set F satisfies the ambiguous monotonicity condition.

To understand why this lemma holds, we claim that type-t_i^* agent i satisfying the strict inequality above can be a whistle-blower: he can unilaterally deviate from the bad equilibrium inducing $f \circ \alpha$ by proposing a profitable reward function h defined by $h(t) = f\big(t_i^*, t_{-i}\big)$ for all $t \in T$. We know that h is a reward function since Lemma 1 tells us that f satisfies ambiguous incentive compatibility. Furthermore, when t_{-i} is unknown, the set of possible realized outcomes under $h\big(\alpha\big(t_i^*, t_{-i}\big)\big) = f\big(t_i^*, \alpha_{-i}\big(t_{-i}\big)\big)$ is a subset of those under $f\big(t_i^*, t_{-i}\big)$. This means that when t_{-i} is unknown, $h\big(\alpha\big(t_i^*, t_{-i}\big)\big)$ is "less uncertain" compared to $f\big(t_i^*, t_{-i}\big)$, and thus $h\big(\alpha\big(t_i^*, t_{-i}\big)\big)$ leads to a weakly higher interim payoff to type-t_i^* agent i than $f\big(t_i^*, t_{-i}\big)$. This, along with the inequality stated in Lemma 2, concludes that the reward function $h\big(\alpha\big(t_i^*, t_{-i}\big)\big)$ can be proposed by t_i^* to dissolve the bad equilibrium inducing $f\big(\alpha\big(t_i^*, t_{-i}\big)\big)$.

Furthermore, we establish an easy-to-check sufficient condition for the bad outcome property in an exchange economy. It is the ambiguous individual rationality condition.

LEMMA 3: *If the social choice set F satisfies the ambiguous individual rationality condition, then it satisfies the bad outcome property.*

We prove in the Appendix that the zero outcome $\mathbf{0} \in \mathbb{R}_+^{nL}$ can serve as a bad outcome. To see this, notice that all agents have nonzero initial endowments. Thus, every ambiguous individually rational social choice function should assign agents nonzero private consumption.

We remark that the exchange economy setup is not needed for Lemmas 1 and 2 to hold. Nonetheless, in an exchange economy, one can explicitly find a bad outcome, the zero outcome, for ambiguous individually rational social choice sets. Hence, we adopt the exchange economy setup to guarantee the bad outcome property. The assumptions of Wald-type maxmin preferences and private value utility functions play a role in the proof of all three lemmas above.

In the Bayesian implementation literature, useful results have been established under private value environments. For example, Palfrey and Srivastava (1989b) have shown that every incentive-compatible social choice function can be fully implemented as the unique undominated Bayesian equilibrium. Such a result provides a simple sufficient condition for full implementation under the private value Bayesian environment. However, their solution concept is different from the one adopted in the current paper. Their solution concept, the undominated Bayesian equilibrium, is a refinement of the Bayesian equilibrium, but ours, the ambiguous equilibrium, is a variant of the Bayesian equilibrium under ambiguous beliefs. Because of this, our results on full implementation under private value environments do not follow as corollaries of their paper.

In the following subsections, we apply our simplified sufficient conditions to verify the implementability of several common solution concepts in exchange economies.

B. *Pareto-Efficient and Individually Rational Social Choice Sets*

We can fully implement the set of all ambiguous Pareto-efficient and individually rational social choice functions in an exchange economy with Wald-type maxmin preferences and private value utility functions.

By applying Theorem 1 and the simplified sufficient conditions in Section IIIA, we have the following result.

COROLLARY 2: *The set of all ambiguous Pareto-efficient and ambiguous individually rational social choice functions F is fully implementable as ambiguous equilibria.*

Corollary 2 is related to the main result of de Castro, Liu, Yannelis (2017b). The main difference is that we look at full implementation of a social choice set, while their paper studies partial implementation of a social choice function.

We remark that Corollary 2 implements the set of ambiguous Pareto-efficient and individually rational social choice functions. This does not mean that every subset of ambiguous Pareto-efficient and individually rational social choice functions is fully implementable because full implementation requires the set of equilibria to coincide with the social choice set. Hence, in the following two subsections, we provide two examples of implementable social choice sets that are ambiguous Pareto efficient and individually rational: the maxmin core and the maxmin value. Their implementability does not follow from Corollary 2.

C. *Maxmin Core and Maxmin Weak Core*

The core of an economy is an important solution concept that is immune to coalitional manipulations. Under the Bayesian framework, Palfrey and Srivastava (1987) have shown that the core notions under asymmetric information are generally not implementable as Bayesian equilibria. The core notions are further studied under the ambiguity aversion framework. See, for example, de Castro, Pesce, and Yannelis (2011) and Moreno-García and Torres-Martínez (2020). In this subsection, we fully implement two core notions under the Wald-type maxmin preferences and private value utility functions.

The first notion is a minor modification of de Castro, Pesce, and Yannelis' (2011) maxmin core allocation. We say a coalition S can block a social choice function under the type profile t^* if by redistributing initial endowment within the coalition, at least one agent $i \in S$ has a higher interim payoff under t_i^*, and every other agent j is not worse off under t_j^*. As a blocking coalition only needs to improve the interim payoff of one member, this maxmin core notion is slightly different from definition 3.12 of de Castro, Pesce, and Yannelis (2011). A maxmin core allocation is unblockable under every type profile $t^* \in T$. In this version of the core notion, agents in a coalition S only propose an alternative social choice function that is feasible given the initial endowments within themselves. They neither share private information with each other nor jointly coordinate on misreporting their types. Hence, each agent $i \in S$ evaluates his interim payoff by considering the worst-case scenario among all possible realizations of $t_{-i} \in T_{-i}$.

DEFINITION 7: *A social choice function f is said to be a **maxmin core** allocation if there does not exist $S \subseteq I$, $t^* \in T$, and a social choice function $y : T \to A_S$, such that*

$$\min_{t_{-i} \in T_{-i}} u_i\big(y\big(t_i^*, t_{-i}\big), t_i^*\big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f\big(t_i^*, t_{-i}\big), t_i^*\big)$$

for all $i \in S$, and the strict inequality holds for some $i \in S$.

One could also consider a weaker core concept by adopting the interim domination notion of Holmström and Myerson (1983). A blocking coalition needs to make one type of some agent better off without hurting any type of any agent in the coalition. As the requirement for such a blocking coalition to exist is rather strong, we call the core concept that is immune to this type of blocking the maxmin weak core.

DEFINITION 8: *A social choice function f is said to be a **maxmin weak core** allocation if there does not exist a coalition $S \subseteq I$ and another social choice function $y : T \to A_S$ such that*

$$\min_{t_{-i} \in T_{-i}} u_i\big(y\big(t\big), t_i\big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f\big(t\big), t_i\big)$$

for all $i \in S$ and $t_i \in T_i$, and the strict inequality holds for some $i \in S$ and $t_i \in T_i$.

The following corollary shows that both of the core concepts are fully implementable.

COROLLARY 3:

(*i*) *The set of all maxmin core allocations is fully implementable as ambiguous equilibria.*

(*ii*) *The set of all maxmin weak core allocations is fully implementable as ambiguous equilibria.*

A key observation in the proof is that both the maxmin core and the maxmin weak core are ambiguous Pareto efficient. This follows from the definitions of the two notions. Moreover, the maxmin core is ambiguous individually rational, which guarantees the bad outcome property. The maxmin weak core may not satisfy the ambiguous individual rationality condition defined in Definition 6, but we demonstrate in the Appendix that the zero outcome can still serve as a bad outcome.

Through the results above, we provide a noncooperative foundation for the maxmin core and the maxmin weak core. It is worth mentioning that theorem 3.1 of Hahn and Yannelis (2001) fully implements the private core as coalitional Bayesian equilibria under the Bayesian framework, although core notions are generally not implementable as Bayesian equilibria. The main difference between our Corollary 3 and their result is that the ambiguous equilibrium adopted in the current paper is a noncooperative game solution concept. It can be interpreted as the Bayesian equilibrium in the ambiguous belief framework. However, their coalitional Bayesian equilibrium can be viewed as a cooperative game solution concept and is a refinement of Bayesian equilibrium. The same difference also exists between our application in Section IIID and their theorem 4.1.

D. *Maxmin Value*

The Shapley value is a widely used solution concept in economic theory. It assigns to each agent his marginal contribution to total surplus in the game. Angelopoulos and Koutsougeras (2015) extend the idea of Shapley value to an asymmetric information economy with Wald-type maxmin preferences. In this subsection, we show that the (interim) maxmin value introduced by them is fully implementable.

For each type profile $t \in T$ and weight profile $\lambda(t) = \left(\lambda_i(t)\right)_{i\in I} \in \mathbb{R}^n_+ \backslash \{\mathbf{0}\}$, define the characteristic function $V_{\lambda,t} : 2^I \to \mathbb{R}$ by $V_{\lambda,t}(\emptyset) = 0$ and

$$V_{\lambda,t}(S) = \max\left\{\sum_{i\in S}\lambda_i(t)\min_{t'_{-i}\in T_{-i}} u_i\left(x\left(t_i,t'_{-i}\right),t_i\right)\middle| x : T \to A_S\right\}$$

for each coalition $S \subseteq I$. The characteristic function measures the coalition's interim worth, i.e., the maximal weighted sum of coalition members' interim payoffs given their initial endowments. From the definition of the characteristic function, it is easy to see that for any disjoint coalitions $S^1, S^2 \subseteq I$, $V_{\lambda,t}\big(S^1 \cup S^2\big) \geq V_{\lambda,t}\big(S^1\big) + V_{\lambda,t}\big(S^2\big)$. This inequality implies that the joint worth of two coalitions is weakly higher than that of having the two coalitions work separately.

The Shapley value of agent i under type profile t is defined as

$$Sh_i\big(V_{\lambda,t}\big) = \sum_{S \ni i} \frac{\big(|S|-1\big)!\big(|I|-|S|\big)!}{|I|!}\Big[V_{\lambda,t}\big(S\big) - V_{\lambda,t}\big(S\backslash\{i\}\big)\Big],$$

which is a way to measure the marginal contribution of agent i, taking into consideration all possible coalitions agent i may join and all possible orders in which members join these coalitions.

The notion below comes from definition 2 of Angelopoulos and Koutsougeras (2015). It requires that each agent's interim payoff from the maxmin value allocation should reflect his marginal contribution.

DEFINITION 9: *A social choice function $f : T \to A$ is a **maxmin value** allocation if for each $t \in T$, there exists a nonzero weight profile $\lambda\big(t\big) \in \mathbb{R}^n_+\backslash\{\mathbf{0}\}$ such that*

$$\lambda_i\big(t\big) \min_{t'_{-i} \in T_{-i}} u_i\Big(f\big(t_i, t'_{-i}\big), t_i\Big) = Sh_i\big(V_{\lambda,t}\big), \quad \forall i \in I.$$

We denote $\lambda\big(t\big) \in \mathbb{R}^n_{++}$ when every dimension of $\lambda\big(t\big)$ is positive. The set of all maxmin value allocations is fully implementable when each maxmin value allocation has weights in $\mathbb{R}^n_{++}$. The proof relies on the fact that a maxmin value allocation is ambiguous Pareto efficient and ambiguous individually rational. The positive weight restriction is used to guarantee ambiguous Pareto efficiency and ambiguous individual rationality.

COROLLARY 4: *Let F be the set of all maxmin value allocations. If for each $f \in F$, its weight profile satisfies $\lambda\big(t\big) \in \mathbb{R}^n_{++}$ for all $t \in T$, then F is fully implementable as ambiguous equilibria.*

The result above provides a noncooperative foundation for the maxmin value.

IV. Conclusion

This paper introduces the maxmin expected utility framework into the problem of fully implementing a social choice set as ambiguous equilibria. This allows us to implement social choice sets that may not be implementable under the Bayesian framework. Under the Wald-type maxmin preferences, we provide easy-to-check sufficient conditions for full implementation, which help us to implement the set of all ambiguous Pareto-efficient and individually rational social choice functions, the maxmin core, the maxmin weak core, and the maxmin value.

APPENDIX A

PROOF OF THEOREM 1:

Part (i).—The proof of this part is along the line of Jackson (1991). For completeness, we include it here. Let F be an implementable social choice set. Thus, there exists a mechanism (M,g) that implements F.

First, to establish the ambiguous incentive compatibility condition for F, we suppose by way of contradiction that there exists $f \in F$, $i \in I$, and two different types $t_i \neq t_i'$ such that type t_i is better off by misreporting t_i'. As $f \in F$, there exists an ambiguous equilibrium σ such that $g\big(\sigma(t)\big) = f(t)$ for all $t \in T$. Define a constant strategy $\bar{\sigma}_i$ for agent i by $\bar{\sigma}_i(t_i'') = \sigma_i(t_i')$ for all $t_i'' \in T_i$. Notice that $g\big(\bar{\sigma}_i(t_i),\sigma_{-i}(t_{-i})\big) = g\big(\sigma_i(t_i'),\sigma_{-i}(t_{-i})\big) = f(t_i',t_{-i})$ for all $t_{-i} \in T_{-i}$. Strategy $\bar{\sigma}_i$ is profitable for type-t_i agent i, contradicting the fact that σ is an ambiguous equilibrium. Hence, F is ambiguous incentive compatible.

Second, to establish the ambiguous monotonicity condition for F, we fix a social choice function $f \in F$ and an unacceptable deception profile α. As $f \in F$ and $f \circ \alpha \notin F$, there exists an ambiguous equilibrium σ such that $\sigma \circ \alpha$ is not an equilibrium. Hence, there exists a type-t_i agent i and a strategy σ_i' such that $\sigma_i'(t_i)$ is a profitable deviation from $\sigma_i \circ \alpha_i$ when other agents follow $\sigma_{-i} \circ \alpha_{-i}$. Define a constant strategy $\bar{\sigma}_i(t_i') = \sigma_i'(t_i)$ for all $t_i' \in T_i$ and a social choice function h by $h(t') = g\big(\bar{\sigma}_i(t_i'),\sigma_{-i}(t_{-i}')\big)$ for all $t' \in T$. Notice that agent i's type does not affect the outcome assigned by h as $\bar{\sigma}_i$ is constant. As $\bar{\sigma}_i$ cannot be a profitable deviation from σ_i when other agents adopt σ_{-i}, we can conclude that $h \in H^f_{\alpha_i(t_i)}$. As $\bar{\sigma}_i$ is a profitable deviation from $\sigma_i \circ \alpha_i$ for type t_i when other agents adopt $\sigma_{-i} \circ \alpha_{-i}$, we can conclude that $h \circ \alpha$ gives type t_i a higher maxmin expected utility than $f \circ \alpha$. Hence, we have established the ambiguous monotonicity condition for F.

Part (ii).—To prove the second part of the theorem, we construct a mechanism (M,g) to implement a social choice set F that satisfies ambiguous incentive compatibility, ambiguous monotonicity, and the bad outcome property.

Each agent i reports a message $m_i = \big(m_i^1, m_i^2, m_i^3, m_i^4, m_i^5\big)$, where $m_i^1 \in T_i$, $m_i^2 \in F$, $m_i^3 \in \mathbb{N}_+$, $m_i^4 \in \mathbb{N}_+$, and m_i^5 is a function from T to A. We partition the message space into M^1, M^2, and M^3 as follows:

$$M^1 = \big\{m \in M \,|\, \exists f \in F \text{ such that } m_i^2 = f \text{ and } m_i^3 = 0,\ \forall i \in I\big\},$$

$$M^2 = \Big\{m \in M \,\Big|\, \exists f \in F,\ i \in I,\ h \in H^f_{m_i^1} \text{ such that } m_i^3 > 0,\ m_i^5 = h,\ m_j^2 = f,\ m_j^3 = 0,\ \forall j \neq i\Big\},$$

$$M^3 = M \backslash \big\{M^1 \cup M^2\big\}.$$

As the bad outcome property holds for F, there exists $\underline{a} \in A$ and $\delta > 0$ such that $u_i\big(f(t'),(t_i,t_{-i}'')\big) - u_i(\underline{a},t) \geq \delta$ for all $i \in I$, $f \in F$, $t_i, t_i' \in T_i$, and

$t_{-i}, t'_{-i}, t''_{-i} \in T_{-i}$. Fix an arbitrary $f^0 \in F$ for the remainder of this proof. Define a social choice function $\underline{a}_\epsilon : T \to A$ where for each $t \in T$,

$$\underline{a}_\epsilon(t) = \begin{cases} f^0(t) & \text{with probability } \epsilon \\ \underline{a} & \text{with probability } 1-\epsilon. \end{cases}$$

In particular, ϵ is sufficiently small and satisfies that

$$\text{(A1)} \qquad \begin{aligned} u_i\big(\underline{a}_\epsilon(t'), t\big) &= \epsilon\, u_i\big(f^0(t'), t\big) + (1-\epsilon)\, u_i(\underline{a}, t) \\ &< u_i(\underline{a}, t) + \delta \le u_i\big(f(t''), (t_i, t'''_{-i})\big) \end{aligned}$$

for all $f \in F$, $i \in I$, $t_i, t'_i, t''_i \in T_i$ and $t_{-i}, t'_{-i}, t''_{-i}, t'''_{-i} \in T_{-i}$. Notice that the "=" follows from the expected utility of the lottery. The "<" relies on the continuity of the expected utility of the lottery with respect to ϵ and the fact that $\delta > 0$. The "$\le$" is a result of the bad outcome property.

By the bad outcome property and the construction of the social choice function $\underline{a}_\epsilon$, we also know that

$$\text{(A2)} \qquad \begin{aligned} u_i\big(\underline{a}_\epsilon(t'), t\big) &= \epsilon\, u_i\big(f^0(t'), t\big) + (1-\epsilon)\, u_i(\underline{a}, t) \\ &\ge \epsilon\big(u_i(\underline{a}, t) + \delta\big) + (1-\epsilon)\, u_i(\underline{a}, t) = u_i(\underline{a}, t) + \epsilon\delta \end{aligned}$$

for all $t, t' \in T$ and $i \in I$. Hence, the social choice function $\underline{a}_\epsilon$ always delivers a higher payoff to all agents than the bad outcome $\underline{a}$.

Let m^1 denote $\big(m_i^1\big)_{i \in I}$. Consider any lottery between $\underline{a}$ and $\underline{a}_\epsilon\big(m^1\big)$. From expression (A1) and the bad outcome property, we know

$$\text{(A3)} \qquad \max_{t_{-i} \in T_{-i}, t' \in T} \Big\{ \alpha\, u_i(\underline{a}, t) + (1-\alpha)\, u_i\big(\underline{a}_\epsilon(t'), t\big) \Big\} < \min_{t_{-i} \in T_{-i}, t'' \in T} u_i\big(f(t''), t\big)$$

for all $f \in F$, $i \in I$, $t_i \in T_i$, and $\alpha \in [0,1]$. This means that for any type-t_i agent $i \in I$, every lottery between $\underline{a}$ and $\underline{a}_\epsilon\big(m^1\big)$ achieves a lower best-case ex post payoff than the worst-case payoff from any social choice outcome. Hence, we call a lottery between $\underline{a}$ and $\underline{a}_\epsilon\big(m^1\big)$ a "bad lottery." Other things equal, we know from expression (A2) that the lower weight the bad lottery imposes on $\underline{a}$, the better the lottery is.

Now we formally define the outcome function g of the mechanism.

If $m \in M^1$, let the outcome be $g(m) = f\big(m^1\big)$, where f is the second component of every agent's message.

If $m \in M^2$, let i be the agent who reports $m_i^3 > 0$. Denote m_i^5 by h. Let $g(m)$ be a lottery whose realization is $h\big(m^1\big)$ with probability $m_i^3 / \big(1 + m_i^3\big)$, $\underline{a}_\epsilon\big(m^1\big)$ with probability $\big(\sum_{j \in I} m_j^4\big) / \Big[\big(1 + m_i^3\big)\big(1 + \sum_{j \in I} m_j^4\big)\Big]$, and $\underline{a}$ with probability $1 / \Big[\big(1 + m_i^3\big)\big(1 + \sum_{j \in I} m_j^4\big)\Big]$.

If $m \in M^3$, let $g(m)$ be a lottery whose realization is $\underline{a}_\epsilon(m^1)$ with probability $\left(\sum_{j\in I} m_j^4\right)/\left(1+\sum_{j\in I} m_j^4\right)$ and $\underline{a}$ with probability $1/\left(1+\sum_{j\in I} m_j^4\right)$.

For each agent $i \in I$, type $t_i \in T_i$, strategy $\sigma_i : T_i \to M_i$, and number $k = 1,\ldots,5$, let $\sigma_i^k(t_i)$ denote the kth component of the message $\sigma_i(t_i)$. Hence, we can decompose σ_i into $\sigma_i = \left(\sigma_i^1, \sigma_i^2, \sigma_i^3, \sigma_i^4, \sigma_i^5\right)$.

To establish that the above mechanism fully implements F, we first fix any $f \in F$ and have the following claim.

CLAIM 1: *A strategy profile* σ^* *satisfying* $\sigma_i^{*1}(t_i) = t_i$, $\sigma_i^{*2}(t_i) = f$, *and* $\sigma_i^{*3}(t_i) = 0$ *for all* $i \in I$ *and* $t_i \in T_i$ *is an ambiguous equilibrium of* (M, g).

PROOF:

Notice that $\sigma^*(t) \in M^1$ and $g\left(\sigma^*(t)\right) = f(t)$ for all $t \in T$. Fix any agent $i \in I$ and type $t_i \in T_i$ for the remainder of the proof of this claim. Consider any alternative strategy σ_i'. The subsequent discussion shows that σ_i' is not a profitable deviation from σ_i^* for t_i.

Case 1: Suppose $\sigma_i'(t_i)$ satisfies that $\sigma_i'^2(t_i) = f$ and $\sigma_i'^3(t_i) = 0$. The new message $\left(\sigma_i'(t_i), \sigma_{-i}^*(t_{-i})\right)$ stays in M^1 for all $t_{-i} \in T_{-i}$. By the ambiguous incentive compatibility condition, the deviation is not profitable for t_i.

Case 2: Suppose that $\sigma_i'(t_i)$ satisfies $\sigma_i'^3(t_i) > 0$ and $\sigma_i'^5(t_i) \in H^f_{\sigma_i'^1(t_i)}$. We denote $\sigma_i'^5(t_i)$ by h. The new message $\left(\sigma_i'(t_i), \sigma_{-i}^*(t_{-i})\right)$ falls in M^2 for all $t_{-i} \in T_{-i}$. For each $t_{-i} \in T_{-i}$, the outcome under type profile t is a lottery of realization $h\left(\sigma_i'^1(t_i), t_{-i}\right)$ with probability $\sigma_i'^3(t_i)/\left(1+\sigma_i'^3(t_i)\right)$, of realization $\underline{a}_\epsilon\left(\sigma_i'^1(t_i), t_{-i}\right)$ with probability $\left(\sigma_i'^4(t_i) + \sum_{j\neq i}\sigma_j^{*4}(t_j)\right)/\left[\left(1+\sigma_i'^3(t_i)\right)\left(1+\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)\right)\right]$, and of realization $\underline{a}$ with probability $1/\left[\left(1+\sigma_i'^3(t_i)\right)\times\left(1+\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)\right)\right]$.

Let $\pi_i^h(t_i) \in \Pi_i(t_i)$ and $\pi_i^f(t_i) \in \Pi_i(t_i)$ satisfy

$$\sum_{t_{-i}\in T_{-i}} u_i\left(h\left(\sigma_i'^1(t_i), t_{-i}\right), t\right)\pi_i^h(t_i)\left[t_{-i}\right] = \min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\left(h\left(\sigma_i'^1(t_i), t_{-i}\right), t\right)\pi_i(t_i)\left[t_{-i}\right]$$

and

$$\text{(A4)} \qquad \sum_{t_{-i}\in T_{-i}} u_i\left(f(t), t\right)\pi_i^f(t_i)\left[t_{-i}\right] = \min_{\pi_i(t_i)\in\Pi_i(t_i)} \sum_{t_{-i}\in T_{-i}} u_i\left(f(t), t\right)\pi_i(t_i)\left[t_{-i}\right],$$

respectively. As $h \in H^f_{\sigma_i'^1(t_i)}$, for this type-$t_i$ agent i,

$$\text{(A5)} \qquad \sum_{t_{-i}\in T_{-i}} u_i\left(h\left(\sigma_i'^1(t_i), t_{-i}\right), t\right)\pi_i^h(t_i)\left[t_{-i}\right] \leq \sum_{t_{-i}\in T_{-i}} u_i\left(f(t), t\right)\pi_i^f(t_i)\left[t_{-i}\right].$$

In addition, expression (A3) implies that

$$\text{(A6)}\qquad \sum_{t_{-i}\in T_{-i}}\left[\frac{\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)}{1+\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)}u_i\left(\underline{a}_\epsilon\left(\sigma_i'^1(t_i),t_{-i}\right),t\right)\right.$$

$$\left.+\frac{1}{1+\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)}u_i\left(\underline{a},t\right)\right]\pi_i^h(t_i)\left[t_{-i}\right]$$

$$<\sum_{t_{-i}\in T_{-i}}u_i\left(f(t),t\right)\pi_i^f(t_i)\left[t_{-i}\right].$$

Hence, we further know that

$$\min_{\pi_i(t_i)\in\Pi_i(t_i)}\sum_{t_{-i}\in T_{-i}}\left[u_i\left(g\left(\sigma_i'(t_i),\sigma_{-i}^*(t_{-i})\right),t\right)\right]\pi_i(t_i)\left[t_{-i}\right]$$

$$\leq\sum_{t_{-i}\in T_{-i}}\left[\frac{\sigma_i'^3(t_i)}{1+\sigma_i'^3(t_i)}u_i\left(h\left(\sigma_i'^1(t_i),t_{-i}\right),t\right)\right.$$

$$+\frac{\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)}{\left(1+\sigma_i'^3(t_i)\right)\left(1+\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)\right)}u_i\left(\underline{a}_\epsilon\left(\sigma_i'^1(t_i),t_{-i}\right),t\right)$$

$$\left.+\frac{1}{\left(1+\sigma_i'^3(t_i)\right)\left(1+\sigma_i'^4(t_i)+\sum_{j\neq i}\sigma_j^{*4}(t_j)\right)}u_i\left(\underline{a},t\right)\right]\pi_i^h(t_i)[t_{-i}]$$

$$<\sum_{t_{-i}\in T_{-i}}u_i\left(f(t_i,t_{-i}),t\right)\pi_i^f(t_i)\left[t_{-i}\right]=\min_{\pi_i(t_i)\in\Pi_i(t_i)}\sum_{t_{-i}\in T_{-i}}u_i\left(f(t_i,t_{-i}),t\right)\pi_i(t_i)\left[t_{-i}\right],$$

where the first inequality follows from the fact that $\pi_i^h(t_i)\in\Pi_i(t_i)$, the second inequality follows from expressions (A5) and (A6), and the equality comes from expression (A4). Hence, we know that such a deviation σ_i' is not profitable for type-t_i agent i.

Case 3: Otherwise, the new message $\left(\sigma_i'(t_i),\sigma_{-i}^*(t_{-i})\right)$ falls in M^3 for all $t_{-i}\in T_{-i}$. By expression (A3), it is easy to see that the deviation is not profitable for type-t_i agent i.

This completes the proof of the claim. ∎

As $g\left(\sigma^*(t)\right)=f(t)$ for all $t\in T$, the first part of the implementability of F has been established.

Now we fix any ambiguous equilibrium σ of the mechanism (M,g) for the remainder of the proof of the theorem. Define a new strategy by $\sigma_i'(t_i)=\left(\sigma_i^1(t_i),\sigma_i^2(t_i),\sigma_i^3(t_i),1+\sigma_i^4(t_i),\sigma_i^5(t_i)\right)$ for each $i\in I$ and $t_i\in T_i$. Also, for each type-t_i agent i, we partition T_{-i} into subsets $T_{-i}^1(t_i)$, $T_{-i}^2(t_i)$, and $T_{-i}^3(t_i)$, where $T_{-i}^k(t_i)=\left\{t_{-i}\in T_{-i}|\sigma(t_i,t_{-i})\in M^k\right\}$ for $k=1,2,3$. Below, we use three

claims to establish that the ambiguous equilibrium σ must lead to an outcome that is consistent with F.

CLAIM 2: *For all* $i \in I$ *and* $t_i \in T_i$, $\sigma_i^3(t_i) = 0$.

PROOF:

Suppose there exists $i \in I$ and $t_i^* \in T_i$ such that $\sigma_i^3(t_i^*) > 0$. We fix this i and t_i^* for the remainder of the proof of the claim. As $\sigma(t_i^*, t_{-i}) \in M^2 \cup M^3$ for all $t_{-i} \in T_{-i}$, we have $T_{-i} = T_{-i}^2(t_i^*) \cup T_{-i}^3(t_i^*)$.

Suppose type-t_i^* agent i deviates with the strategy σ_i'. The deviation will impose a smaller weight on $\underline{a}$ at each state (t_i^*, t_{-i}) for all $t_{-i} \in T_{-i}$. Hence, the increase in type-t_i^* agent i's expected utility by adopting σ_i' instead of σ_i under belief $\pi_i(t_i^*) \in \Pi_i(t_i^*)$, denoted by $\Delta(\pi_i(t_i^*))$, is equal to

$$\frac{1}{1+\sigma_i^3(t_i^*)} \sum_{t_{-i} \in T_{-i}^2(t_i^*)} \frac{u_i\big(\underline{a}_\epsilon(\sigma^1(t_i^*, t_{-i})), (t_i^*, t_{-i})\big) - u_i\big(\underline{a}, (t_i^*, t_{-i})\big)}{\big(2+\sigma_i^4(t_i^*)+\sum_{j\neq i}\sigma_j^4(t_j)\big)\big(1+\sigma_i^4(t_i^*)+\sum_{j\neq i}\sigma_j^4(t_j)\big)} \pi_i(t_i^*)[t_{-i}]$$

$$+ \sum_{t_{-i} \in T_{-i}^3(t_i^*)} \frac{u_i\big(\underline{a}_\epsilon(\sigma^1(t_i^*, t_{-i})), (t_i^*, t_{-i})\big) - u_i\big(\underline{a}, (t_i^*, t_{-i})\big)}{\big(2+\sigma_i^4(t_i^*)+\sum_{j\neq i}\sigma_j^4(t_j)\big)\big(1+\sigma_i^4(t_i^*)+\sum_{j\neq i}\sigma_j^4(t_j)\big)} \pi_i(t_i^*)[t_{-i}].$$

Notice that there exists $t_{-i} \in T_{-i} = T_{-i}^2(t_i^*) \cup T_{-i}^3(t_i^*)$ such that $\pi_i(t_i^*)[t_{-i}] > 0$. This, along with expression (A2), demonstrates that $\Delta(\pi_i(t_i^*)) > 0$.

After deviation, the maxmin expected utility for type-t_i^* agent i is

$$\min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \left\{ \sum_{t_{-i} \in T_{-i}} u_i\big(g(\sigma(t_i^*, t_{-i})), (t_i^*, t_{-i})\big) \pi_i(t_i^*)[t_{-i}] + \Delta(\pi_i(t_i^*)) \right\}$$
$$> \min_{\pi_i(t_i^*) \in \Pi_i(t_i^*)} \sum_{t_{-i} \in T_{-i}} \Big[u_i\big(g(\sigma(t_i^*, t_{-i})), (t_i^*, t_{-i})\big) \Big] \pi_i(t_i^*)[t_{-i}],$$

where the strict inequality comes from the compactness of $\Pi_i(t_i^*)$ and the fact that $\Delta(\pi_i(t_i^*)) > 0$ for all $\pi_i(t_i^*) \in \Pi_i(t_i^*)$. Hence, type-$t_i^*$ agent i is better off by deviating, contradicting the fact that σ is an ambiguous equilibrium. ∎

CLAIM 3: *For all* $t \in T$, $\sigma(t) \in M^1$.

PROOF:

For each $t \in T$, Claim 2 implies that $\sigma(t) \notin M^2$, and thus we only need to show that $\sigma(t) \notin M^3$ to establish Claim 3. Suppose by way of contradiction that there exists $t^* \in T$ such that $\sigma(t^*) \in M^3$. We fix this t^* and an arbitrary agent $i \in I$ for the remainder of the proof of the claim. Notice that $t_{-i}^* \in T_{-i}^3(t_i^*)$.

Messages $\sigma_i'(t_i^*)$ and $\sigma_i(t_i^*)$ only differ in their fourth components. From the design of (M, g), we have the following two observations. First, for each $t_{-i} \in T_{-i}^1(t_i^*)$, $g(\sigma_i'(t_i^*), \sigma_{-i}(t_{-i})) = g(\sigma_i(t_i^*), \sigma_{-i}(t_{-i}))$. Second, for each $t_{-i} \in T_{-i}^3(t_i^*)$, $(\sigma_i'(t_i^*), \sigma_{-i}(t_{-i})) \in M^3$. The only difference between lotteries $g(\sigma_i'(t_i^*), \sigma_{-i}(t_{-i}))$

and $g\big(\sigma(t_i^*, t_{-i})\big)$ is that the bad outcome $\underline{a}$ is realized with lower probability under $g\big(\sigma_i'(t_i^*), \sigma_{-i}(t_{-i})\big)$.

Suppose Assumption 1 holds for t_i^*. Under each $\pi_i(t_i^*) \in \Pi_i(t_i^*)$, the increase in type-$t_i^*$ agent i's expected utility by adopting σ_i' instead of σ_i, denoted by $\Delta\big(\pi_i(t_i^*)\big)$, is

$$\sum_{t_{-i} \in T_{-i}^3(t_i^*)} \frac{u_i\Big(\underline{a}_\epsilon\big(\sigma^1(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big) - u_i\big(\underline{a}, (t_i^*, t_{-i})\big)}{\Big(2 + \sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)\Big)\Big(1 + \sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)\Big)} \pi_i(t_i^*)[t_{-i}] \; > \; 0.$$

The strict inequality above uses the assumption that $\pi_i(t_i^*)[t_{-i}] > 0$ for all $t_{-i} \in T_{-i}$, the fact that $t_{-i}^* \in T_{-i}^3(t_i^*)$, and expression (A2). Thus, by an argument that is similar to Claim 2, it is profitable for type-t_i^* agent i to deviate from $\sigma_i(t_i^*)$ to $\sigma_i'(t_i^*)$, which contradicts the fact that σ is an ambiguous equilibrium.

Suppose Assumption 2 holds for t_i^* instead. Thus, type t_i^* has Wald-type maxmin preference. Thus, by deviating with σ_i', type-t_i^* agent i has a payoff of

$$\begin{aligned}
&\min_{t_{-i} \in T_{-i}} u_i\Big(g\big(\sigma_i'(t_i^*), \sigma_{-i}(t_{-i})\big), (t_i^*, t_{-i})\Big) \\
&\quad = \min_{t_{-i} \in T_{-i}^3(t_i^*)} \left\{ \frac{1 + \sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)}{2 + \sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)} u_i\Big(\underline{a}_\epsilon\big(\sigma^1(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big) \right. \\
&\qquad\qquad \left. + \frac{1}{2 + \sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)} u_i\big(\underline{a}, (t_i^*, t_{-i})\big) \right\} \\
&\quad > \min_{t_{-i} \in T_{-i}^3(t_i^*)} \left\{ \frac{\sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)}{1 + \sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)} u_i\Big(\underline{a}_\epsilon\big(\sigma^1(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big) \right. \\
&\qquad\qquad \left. + \frac{1}{1 + \sigma_i^4(t_i^*) + \sum_{j \neq i} \sigma_j^4(t_j)} u_i\big(\underline{a}, (t_i^*, t_{-i})\big) \right\} \\
&\quad = \min_{t_{-i} \in T_{-i}} u_i\Big(g\big(\sigma(t_i^*, t_{-i})\big), (t_i^*, t_{-i})\Big).
\end{aligned}$$

To see where the two equalities come from, first notice that $T_{-i}^3(t_i^*) \neq \emptyset$ as $t_{-i}^* \in T_{-i}^3(t_i^*)$. Also, recall the definition of g on M^1 and M^3 as well as expression (A3). Hence, the minimums in the first and last rows are attained by type profiles $t_{-i} \in T_{-i}^3(t_i^*)$. The strict inequality comes from the decreased weight on $\underline{a}$ under the σ_i', expression (A2), and compactness of $T_{-i}^3(t_i^*)$. Hence, $\sigma_i'(t_i^*)$ is more profitable than $\sigma_i(t_i^*)$, contradicting the fact that σ is an ambiguous equilibrium.

This completes the proof of the claim. ∎

CLAIM 4: *There exists* $f' \in F$ *such that* $g\big(\sigma(t)\big) = f'(t)$ *for all* $t \in T$.

PROOF:

From the previous claim, there exists $f \in F$ such that $g\big(\sigma(t)\big) = f\big(\sigma^1(t)\big)$ for all $t \in T$. Suppose by way of contradiction that there does not exist $f' \in F$ such that $g\big(\sigma(t)\big) = f'(t)$ for all $t \in T$. Define a deception profile α by $\alpha_i(t_i) = \sigma_i^1(t_i)$ for all $i \in I$ and $t_i \in T_i$. Given the social choice set F and the social choice function $f \in F$, the deception profile α is unacceptable. By ambiguous monotonicity of F, there exists $i \in I$, $t_i^* \in T_i$, and $h \in H^f_{\alpha_i(t_i^*)}$ such that

$$\text{(A7)} \qquad \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\alpha(t_i^*,t_{-i})\big),(t_i^*,t_{-i})\Big)\,\pi_i(t_i^*)\big[t_{-i}\big] > \min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(f\big(\alpha(t_i^*,t_{-i})\big),(t_i^*,t_{-i})\Big)\,\pi_i(t_i^*)\big[t_{-i}\big].$$

Fix this agent i, type t_i^*, and reward function h for the rest of the proof.

For agent i, define a strategy σ_i'' by $\sigma_i''(t_i^*) = \big(\sigma_i^1(t_i^*),\sigma_i^2(t_i^*),K^*,\sigma_i^4(t_i^*),h\big)$ and $\sigma_i''(t_i) = \sigma_i(t_i)$ for $t_i \neq t_i^*$, where $K^* > 0$ is a large integer. Thus, $\big(\sigma_i''(t_i^*),\sigma_{-i}(t_{-i})\big) \in M^2$ for all $t_{-i} \in T_{-i}$. By deviating, type-t_i^* agent i has an interim payoff of

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \Bigg\{ \sum_{t_{-i}\in T_{-i}} \Bigg[\frac{K^*}{1+K^*} u_i\Big(h\big(\alpha(t_i^*,t_{-i})\big),(t_i^*,t_{-i})\Big) + \frac{1}{\big(1+K^*\big)\big(1+\sigma_i^4(t_i^*)+\sum_{j\neq i}\sigma_j^4(t_j)\big)} u_i\big(\underline{a},(t_i^*,t_{-i})\big) + \frac{\sigma_i^4(t_i^*)+\sum_{j\neq i}\sigma_j^4(t_j)}{\big(1+K^*\big)\big(1+\sigma_i^4(t_i^*)+\sum_{j\neq i}\sigma_j^4(t_j)\big)} \times u_i\Big(\underline{a}_\epsilon\big(\alpha(t_i^*,t_{-i})\big),(t_i^*,t_{-i})\Big)\Bigg]\pi_i(t_i^*)\big[t_{-i}\big]\Bigg\}.$$

When K^* is sufficiently large, the value above is sufficiently close to

$$\min_{\pi_i(t_i^*)\in\Pi_i(t_i^*)} \sum_{t_{-i}\in T_{-i}} u_i\Big(h\big(\alpha(t_i^*,t_{-i})\big),(t_i^*,t_{-i})\Big)\,\pi_i(t_i^*)\big[t_{-i}\big].$$

By expression (A7), this deviation is profitable for type-t_i^* agent i with a large K^*, a contradiction.

This completes the proof of the claim. ∎

In view of the four claims, we have established that (M,g) implements F. ∎

PROOF OF COROLLARY 1:

To establish this corollary, we adopt the same mechanism as in the Proof of Theorem 1 and replace F with the singleton $\{f\}$. For any ambiguous equilibrium

σ, after establishing Claim 2, it is immediate that $\sigma_i^2(t_i) = f$ and $\sigma_i^3(t_i) = 0$ for all $i \in I$ and $t_i \in T_i$. Namely, we can claim that $\sigma(t) \in M^1$ for all $t \in T$ without going through the Proof of Claim 3. The rest of the argument follows from Theorem 1. Assumptions 1 and 2 are not used in the proof when $F = \{f\}$. ∎

PROOF OF LEMMA 1:

The proof comes from proposition A.1 of de Castro and Yannelis (2018) and is omitted here. ∎

PROOF OF LEMMA 2:

Suppose for any $f \in F$ and unacceptable deception profile α, there exists $i \in I$ and $t_i^* \in T_i$ satisfying

$$\min_{t_{-i} \in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i^*\big) > \min_{t_{-i} \in T_{-i}} u_i\Big(f\big(\alpha(t_i^*, t_{-i})\big), t_i^*\Big).$$

Let a social choice function h be defined as $h(t) = f(t_i^*, t_{-i})$ for all $t \in T$. Define $H(t_i^*) = \{a \in A \mid \exists t_{-i} \in T_{-i} \text{ such that } a = h(t_i^*, t_{-i})\}$ and $H(t_i^*, \alpha) = \{a \in A \mid \exists t_{-i} \in T_{-i} \text{ such that } a = h\big(\alpha(t_i^*, t_{-i})\big)\}$. It is easy to show that $H(t_i^*, \alpha) \subseteq H(t_i^*)$. To see this, for any $a \in H(t_i^*, \alpha)$, there exists $t_{-i} \in T_{-i}$ such that $a = h\big(\alpha(t_i^*, t_{-i})\big) = f\big(t_i^*, \alpha_{-i}(t_{-i})\big) = h\big(t_i^*, \alpha_{-i}(t_{-i})\big) \in H(t_i^*)$. Thus, we know

$$\min_{t_{-i} \in T_{-i}} u_i\Big(h\big(\alpha(t_i^*, t_{-i})\big), t_i^*\Big) = \min_{a \in H(t_i^*, \alpha)} u_i(a, t_i^*) \geq \min_{a \in H(t_i^*)} u_i(a, t_i^*)$$

$$= \min_{t_{-i} \in T_{-i}} u_i\big(h(t_i^*, t_{-i}), t_i^*\big) = \min_{t_{-i} \in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i^*\big) > \min_{t_{-i} \in T_{-i}} u_i\Big(f\big(\alpha(t_i^*, t_{-i})\big), t_i^*\Big).$$

In the expression above, the first two equalities follow from the definition of $H(t_i^*, \alpha)$ and $H(t_i^*)$, the Wald-type maxmin preferences, and the private value utility functions. The third equality follows from the definition of h. The weak inequality comes from the fact that $H(t_i^*, \alpha) \subseteq H(t_i^*)$. The strict inequality comes from the supposition.

The argument below shows that $h \in H^f_{\alpha_i(t_i^*)}$. As f is ambiguous incentive compatible, we know

$$\min_{t_{-i} \in T_{-i}} u_i\big(f(t), t_i\big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f(t_i^*, t_{-i}), t_i\big)$$

for all $t_i \in T_i$. Also, since $h(t_i, t_{-i}) = f(t_i^*, t_{-i})$ for all $t \in T$, by replacing $f(t_i^*, t_{-i})$ with $h\big(\alpha_i(t_i^*), t_{-i}\big)$ on the right-hand side of the weak inequality above, one has established that $h \in H^f_{\alpha_i(t_i^*)}$.

Hence, we have proved the lemma. ∎

PROOF OF LEMMA 3:

For each $i \in I$ and $t_i \in T_i$, define a set $A_{i,t_i} \subseteq A$ in the following way:

$$A_{i,t_i} = \{a \in A | u_i(a,t_i) \geq u_i(e,t_i)\}.$$

Notice that the set A_{i,t_i} is a compact set since u_i is continuous in outcomes due to the expected utility setup and A is compact in this exchange economy. Define $\underline{A} = \bigcup_{t\in T}\bigcap_{i\in I} A_{i,t_i}$, which is also a compact set due to the finiteness of I and T.

We claim that for each $a \in \underline{A}$, $i \in I$, and $t_i \in T_i$, it must be true that $u_i(a,t_i) > u_i(\mathbf{0},t_i)$, where $\mathbf{0} \in \mathbb{R}^{nL}_+$. Otherwise, there exists $a \in \underline{A}$, $i \in I$, and $t_i \in T_i$, such that $u_i(a,t_i) \leq u_i(\mathbf{0},t_i)$. This inequality, along with the fact that u_i is monotone, implies that a gives agent i zero consumption at almost all realizations. This contradicts the fact that $a \in \underline{A}$, as e is a nonzero pure outcome.

Notice that the sets $\underline{A}$, I, and T_i are compact. Thus, δ below is well defined and positive:

$$\delta = \min_{a\in\underline{A}, i\in I, t_i\in T_i} \{u_i(a,t_i) - u_i(\mathbf{0},t_i)\} > 0. \tag{A8}$$

As F satisfies the ambiguous individual rationality condition, by the equivalence between ambiguous individual rationality and ex post individual rationality, we know $\{a \in A \,|\, \exists t' \in T \text{ such that } a = f(t')\} \subseteq \underline{A}$ for each $f \in F$. Thus,

$$\min_{t'\in T} u_i(f(t'),t_i) - u_i(\mathbf{0},t_i) \geq \min_{a\in\underline{A}} u_i(a,t_i) - u_i(\mathbf{0},t_i) \geq \delta \tag{A9}$$

for all $f \in F$, $i \in I$, and $t_i \in T_i$. Notice that the first inequality relies on the fact that $\{a \in A | \exists t' \in T \text{ such that } a = f(t')\} \subseteq \underline{A}$ and that agents have the Wald-type maxmin preferences as well as private valuations. The second inequality follows from expression (A8).

From expression (A9), we know that $\mathbf{0} \in \mathbb{R}^{nL}_+$ is a "bad outcome." The bad outcome property thus holds for F. ∎

PROOF OF COROLLARY 2:

Since each $f \in F$ is ambiguous Pareto efficient, F satisfies the ambiguous incentive compatibility condition by Lemma 1.

Given the social choice set F and any social choice function $f \in F$, if α is unacceptable, then $f \circ \alpha$ is not ambiguous Pareto efficient or not ambiguous individually rational. We claim that there must exist an agent $i \in I$ and a type $t_i^* \in T_i$ such that

$$\min_{t_{-i}\in T_{-i}} u_i(f(t_i^*,t_{-i}),t_i^*) > \min_{t_{-i}\in T_{-i}} u_i(f(\alpha(t_i^*,t_{-i})),t_i^*).$$

Otherwise, since $f \circ \alpha$ is at least as good as $f \in F$ for every agent under every type, $f \circ \alpha$ should also be ambiguous Pareto efficient and ambiguous individually

rational. This would contradict the supposition that $f \circ \alpha \notin F$. Hence, we know from Lemma 2 that F satisfies the ambiguous monotonicity condition.

As F satisfies ambiguous individual rationality, the bad outcome property of F holds by Lemma 3.

In view of Theorem 1, F is implementable as ambiguous equilibria. ∎

PROOF OF COROLLARY 3:

Part (i).—Let F be the set of all maxmin core allocations.

By setting $S = I$, it is easy to see that F satisfies the ambiguous Pareto efficiency condition. By Lemma 1, F is ambiguous incentive compatible.

To establish the ambiguous monotonicity condition, we begin with the social choice set F, a social choice function $f \in F$, and an unacceptable deception profile α. As $f \circ \alpha$ is not a maxmin core allocation, there exists $S \subseteq I$, $t^* \in T$, and $y : T \to A_S$ such that

$$\min_{t_{-i} \in T_{-i}} u_i\big(y\big(t_i^*, t_{-i}\big), t_i^*\big) \geq \min_{t_{-i} \in T_{-i}} u_i\Big(f\big(\alpha\big(t_i^*, t_{-i}\big)\big), t_i^*\Big) \tag{A10}$$

for all $i \in S$, and the strict inequality holds for some $i \in S$. Suppose by way of contradiction that

$$\min_{t_{-i} \in T_{-i}} u_i\Big(f\big(\alpha\big(t\big)\big), t_i\Big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f\big(t\big), t_i\big) \quad \forall i \in I,\ t_i \in T_i.$$

From expression (A10), we further know that

$$\min_{t_{-i} \in T_{-i}} u_i\big(y\big(t_i^*, t_{-i}\big), t_i\big) \geq \min_{t_{-i} \in T_{-i}} u_i\big(f\big(t_i^*, t_{-i}\big), t_i^*\big) \tag{A11}$$

for all $i \in S$, and the strict inequality holds for some $i \in S$. This contradicts the fact that $f \in F$, the maxmin core, as $y : T \to A_S$. Hence, we know that there exists an agent $i \in I$ and a type $t_i \in T_i$ such that

$$\min_{t_{-i} \in T_{-i}} u_i\big(f\big(t_i, t_{-i}\big), t_i\big) > \min_{t_{-i} \in T_{-i}} u_i\Big(f\big(\alpha\big(t_i, t_{-i}\big)\big), t_i\Big).$$

By Lemma 2, F satisfies the ambiguous monotonicity condition.

By setting S to be singleton coalitions in Definition 7, we know each $f \in F$ satisfies ambiguous individual rationality. By Lemma 3, F satisfies the bad outcome property.

In view of Theorem 1, F is implementable as ambiguous equilibria.

Part (ii).—Let F be the set of all maxmin weak core allocations.

The condition of ambiguous Pareto efficiency can be proved immediately by setting $S = I$ in Definition 8. The ambiguous incentive compatibility condition follows from Lemma 1. The verification of the ambiguous monotonicity condition is similar to Part (i) except for minor changes, and thus we omit the details.

It remains to verify the bad outcome property for F. We define $\underline{A} = \{a \in A \mid \exists f \in F \text{ and } t' \in T \text{ such that } a = f(t')\}$ and

$$\delta = \inf_{a \in \underline{A}, i \in I, t_i \in T_i} \{u_i(a, t_i) - u_i(\mathbf{0}, t_i)\}.$$

To show that $\mathbf{0} \in \mathbb{R}_+^{nL}$ can serve as a bad outcome, it suffices to prove that $\delta > 0$.

Suppose by way of contradiction that $\delta = 0$. By compactness of I, T_i, and T, there exists $i \in I$, $t_i \in T_i$, $t' \in T$, and a sequence $(f^k \in F)_{k=1,2,\ldots}$ such that

$$\lim_{k \to \infty} u_i(f^k(t'), t_i) = u_i(\mathbf{0}, t_i).$$

By monotonicity and continuity of utility functions, we also have

$$\lim_{k \to \infty} u_i(f^k(t'), t_i') = u_i(\mathbf{0}, t_i') < u_i(e, t_i'), \tag{A12}$$

where the strict inequality comes from the fact that e is nonzero.

For each k, since $f^k \in F$, by setting $S = \{i\}$ in Definition 8, we know it is impossible that e is better than f^k in maxmin expected utility for all types of agent i. Hence, whenever k is sufficiently large such that $\min_{t_{-i} \in T_{-i}} u_i(f^k(t_i', t_{-i}), t_i') \leq u_i(f^k(t'), t_i') < u_i(e, t_i')$, there exists a type $t_i^* \neq t_i'$ such that

$$\min_{t_{-i} \in T_{-i}} u_i(f^k(t_i^*, t_{-i}), t_i^*) \geq u_i(e, t_i^*).$$

The type t_i^* may depend on k. However, as T_i is finite, $(f^k \in F)_{k=1,2,\ldots}$ has a subsequence for which the weak inequality holds for the same t_i^*. Hence, it is without loss of generality to assume that there exists $K_1 > 0$ and $t_i^* \in T_i$ such that

$$u_i(f^k(t_i^*, t_{-i}), t_i^*) \geq \min_{t_{-i} \in T_{-i}} u_i(f^k(t_i^*, t_{-i}), t_i^*) \geq u_i(e, t_i^*), \quad \forall t_{-i} \in T_{-i},\ k \geq K_1.$$

Define $\hat{A} = \{a \in A : u_i(a, t_i^*) \geq u_i(e, t_i^*)\}$, which is compact due to the compactness of A and continuity of utility functions. As $f^k(t_i^*, t_{-i}) \in \hat{A}$ for all $t_{-i} \in T_{-i}$ and $k \geq K_1$, we have

$$u_i(f^k(t_i^*, t_{-i}), t_i') \geq \min_{a \in \hat{A}} u_i(a, t_i') > u_i(\mathbf{0}, t_i'), \quad \forall t_{-i} \in T_{-i},\ k \geq K_1. \tag{A13}$$

The above strict inequality holds because $\hat{A}$ is compact, utility functions are continuous and monotone, and each $a \in \hat{A}$ assigns nonzero private consumption to agent i for a positive measure of realizations.

By expressions (A12) and (A13), we further know there exists k sufficiently large such that

$$u_i(f^k(t_i^*, t_{-i}), t_i') > u_i(f^k(t'), t_i'), \quad \forall t_{-i} \in T_{-i}.$$

Since T_{-i} is finite,

$$\text{(A14)} \quad \min_{t_{-i}\in T_{-i}} u_i\big(f^k(t_i^*,t_{-i}),t_i'\big) > u_i\big(f^k(t'),t_i'\big) \geq \min_{t_{-i}\in T_{-i}} u_i\big(f^k(t_i',t_{-i}),t_i'\big).$$

We fix this k for the remainder of the proof and define an alternative social choice function $y : T \to A$ as follows:

$$y(t) = \begin{cases} f^k(t_i^*,t_{-i}) & \text{if } t_i = t_i' \\ f^k(t) & \text{otherwise.} \end{cases}$$

At last, we claim that y Pareto improves upon f^k. From expression (A14) and the definition of y, we know that y makes type-t_i' agent i better off compared to f^k and does not change the payoff of any type $t_i \neq t_i'$. For all $j \neq i$ and $t_j \in T_j$, define $Y^y(t_j) = \big\{a \in A | \exists t_{-j} \in T_{-j} \text{ such that } y(t) = a\big\}$ and $Y^{f^k}(t_j) = \big\{a \in A | \exists t_{-j} \in T_{-j} \text{ such that } f^k(t) = a\big\}$. It must be the case that $Y^y(t_j) \subseteq Y^{f^k}(t_j)$ for all j and t_j. To see this, notice that for $t \in T$ such that $t_i = t_i'$, $y(t) = f^k(t_i^*,t_{-i}) \in Y^{f^k}(t_j)$, and for other $t \in T$, $y(t) = f^k(t) \in Y^{f^k}(t_j)$. Hence,

$$\begin{aligned} \min_{t_{-j}\in T_{-j}} u_j\big(y(t),t_j\big) &= \min_{a\in Y^y(t_j)} u_j(a,t_j) \\ &\geq \min_{a\in Y^{f^k}(t_j)} u_j(a,t_j) = \min_{t_{-j}\in T_{-j}} u_j\big(f^k(t),t_j\big), \quad \forall j \neq i,\ t_j \in T_j. \end{aligned}$$

The weak inequality above uses the fact that $Y^y(t_j) \subseteq Y^{f^k}(t_j)$. The equalities above rely on the Wald-type maxmin assumption, the private value assumption on utility functions, and the definitions of $Y^{f^k}(t_j)$ and $Y^y(t_j)$.

To this end, we have established that y Pareto improves upon f^k, contradicting the fact that $f^k \in F$ is ambiguous Pareto efficient. Hence, we have $\delta = 0$. Since the zero outcome is a bad outcome, F satisfies the bad outcome property. ∎

PROOF OF COROLLARY 4:

We first establish that F satisfies the ambiguous Pareto efficiency condition. Suppose not. Thus, there exists $f \in F$ whose weight profile is $\lambda : T \to \mathbb{R}^n_{++}$ and a social choice function $y : T \to A$, such that

$$\min_{t_{-i}\in T_{-i}} u_i\big(y(t),t_i\big) \geq \min_{t_{-i}\in T_{-i}} u_i\big(f(t),t_i\big)$$

for all $i \in I$ and $t_i \in T_i$, and the strict inequality holds for some $i \in I$ and $t_i \in T_i$. Fix an agent j and a type t_j^* for which the strict inequality holds and an arbitrary t_{-j}^* for the remainder of this paragraph. As $\lambda(t^*) \in \mathbb{R}^n_{++}$, we know

$$\begin{aligned} \sum_{i\in I} \lambda_i(t^*) \min_{t_{-i}\in T_{-i}} u_i\big(y(t_i^*,t_{-i}),t_i^*\big) &> \sum_{i\in I} \lambda_i(t^*) \min_{t_{-i}\in T_{-i}} u_i\big(f(t_i^*,t_{-i}),t_i^*\big) \\ &= \sum_{i\in I} Sh_i\big(V_{\lambda,t^*}\big) = V_{\lambda,t^*}(I), \end{aligned}$$

where the last two equalities follow from the definition of maxmin value allocation and Shapley value. As y is feasible, this is a contradiction with the definition of $V_{\lambda,t^*}(I)$. Hence, F is ambiguous Pareto efficient. By Lemma 1, F is also ambiguous incentive compatible.

To establish the ambiguous monotonicity condition, we begin with the social choice set F, a value allocation $f \in F$ whose weight profile is $\lambda : T \to \mathbb{R}^n_{++}$, and an unacceptable deception profile α. We claim that there exists an agent $i \in I$ and a type $t_i^* \in T_i$ such that

$$\min_{t_{-i}\in T_{-i}} u_i\big(f\big(t_i^*, t_{-i}\big), t_i^*\big) > \min_{t_{-i}\in T_{-i}} u_i\big(f\big(\alpha\big(t_i^*, t_{-i}\big)\big), t_i^*\big).$$

To see this, since $f\circ\alpha \notin F$, $f\circ\alpha$ cannot be a maxmin value allocation under the same weight profile $\lambda : T \to \mathbb{R}^n_{++}$. Thus, there must exist some type of an agent to whom f and $f\circ\alpha$ lead to different interim utility levels. Since f is ambiguous Pareto efficient, we know $f\circ\alpha$ has to lower at least one agent's interim utility under some type. Hence, we can obtain the above inequality. By Lemma 2, F satisfies the ambiguous monotonicity condition.

In the end, we verify the bad outcome property. We first establish ambiguous individual rationality. Notice that for each $t \in T$ and $i \in I$,

$$\begin{aligned}
&\lambda_i(t) \min_{t'_{-i}\in T_{-i}} u_i\big(f\big(t_i, t'_{-i}\big), t_i\big) \\
&= Sh_i(V_{\lambda,t}) = \sum_{S\ni i} \frac{(|S|-1)!(|I|-|S|)!}{|I|!}\big[V_{\lambda,t}(S) - V_{\lambda,t}\big(S\backslash\{i\}\big)\big] \\
&\geq \sum_{S\ni i} \frac{(|S|-1)!(|I|-|S|)!}{|I|!}\big[V_{\lambda,t}\big(\{i\}\big) + V_{\lambda,t}\big(S\backslash\{i\}\big) - V_{\lambda,t}\big(S\backslash\{i\}\big)\big] \\
&= \sum_{S\ni i} \frac{(|S|-1)!(|I|-|S|)!}{|I|!} V_{\lambda,t}\big(\{i\}\big) = V_{\lambda,t}\big(\{i\}\big) = \lambda_i(t)\, u_i(e, t_i),
\end{aligned}$$

where the inequality follows from $V_{\lambda,t}(S) \geq V_{\lambda,t}\big(\{i\}\big) + V_{\lambda,t}\big(S\backslash\{i\}\big)$, a property of the characteristic function. As $\lambda_i(t) > 0$, the inequality above implies that

$$\min_{t'_{-i}\in T_{-i}} u_i\big(f\big(t_i, t'_{-i}\big), t_i\big) \geq u_i(e, t_i).$$

Hence, F satisfies the ambiguous individual rationality condition. By Lemma 3, the bad outcome property holds for F.

In view of Theorem 1, F is fully implementable as ambiguous equilibria. ∎

REFERENCES

Angelopoulos, Angelos, and Leonidas C. Koutsougeras. 2015. "Value Allocation under Ambiguity." *Economic Theory* 59 (1): 147–67.

Bergemann, Dirk, and Stephen Morris. 2011. "Robust Implementation in General Mechanisms." *Games and Economic Behavior* 71 (2): 261–81.

Bewley, Truman F. 2002. "Knightian Decision Theory. Part I." *Decisions in Economics and Finance* 25 (2): 79–110.

Bodoh-Creed, Aaron L. 2012. "Ambiguous Beliefs and Mechanism Design." *Games and Economic Behavior* 75 (2): 518–37.

Bose, Subir, and Ludovic Renou. 2014. "Mechanism Design with Ambiguous Communication Devices." *Econometrica* 82 (5): 1853–72.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio. 2011. "Uncertainty Averse Preferences." *Journal of Economic Theory* 146 (4): 1275–1330.

de Castro, Luciano I., Zhiwei Liu, and Nicholas C. Yannelis. 2017a. "Ambiguous Implementation: The Partition Model." *Economic Theory* 63 (1): 233–61.

de Castro, Luciano I., Zhiwei Liu, and Nicholas C. Yannelis. 2017b. "Implementation under Ambiguity." *Games and Economic Behavior* 101: 20–33.

de Castro, Luciano I., Marialaura Pesce, and Nicholas C. Yannelis. 2011. "Core and Equilibria under Ambiguity." *Economic Theory* 48 (2-3): 519-48.

de Castro, Luciano, and Nicholas C. Yannelis. 2018. "Uncertainty, Efficiency and Incentive Compatibility: Ambiguity Solves the Conflict between Efficiency and Incentive Compatibility." *Journal of Economic Theory* 177: 678–707.

di Tillio, Alfredo, Nenad Kos, and Matthias Messner. 2017. "The Design of Ambiguous Mechanisms." *Review of Economic Studies* 84 (1): 237–76.

Dutta, Bhaskar, and Arunava Sen. 1991. "A Necessary and Sufficient Condition for Two-Person Nash Implementation." *Review of Economic Studies* 58 (1): 121–28.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75 (4): 643–69.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with Non-unique Prior." *Journal of Mathematical Economics* 18 (2): 141–53.

Guo, Huiyi. 2019. "Mechanism Design with Ambiguous Transfers: An Analysis in Finite Dimensional Naive Type Spaces." *Journal of Economic Theory* 183: 76–105.

Hahn, Guangsug, and Nicholas C. Yannelis. 2001. "Coalitional Bayesian Nash Implementation in Differential Information Economies." *Economic Theory* 18 (2): 485–509.

Holmström, Bengt, and Roger B. Myerson. 1983. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* 6 (51): 1799–1819.

Jackson, Mathew O. 1991. "Bayesian Implementation." *Econometrica* 59 (2): 461–77.

Liu, Zhiwei. 2016. "Implementation of Maximin Rational Expectations Equilibrium." *Economic Theory* 62 (4): 813–37.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006. "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences." *Econometrica* 74 (6): 1447–98.

Maskin, Eric. 1999. "Nash Equilibrium and Welfare Optimality." *Review of Economic Studies* 66 (1): 23–38.

Moreno-García, Emma, and Juan Pablo Torres-Martínez. 2020. "Information within Coalitions: Risk and Ambiguity." *Economic Theory* 69 (1): 125–47.

Palfrey, Thomas R., and Sanjay Srivastava. 1987. "On Bayesian Implementable Allocations." *Review of Economic Studies* 54 (2): 193–208.

Palfrey, Thomas R., and Sanjay Srivastava. 1989a. "Implementation with Incomplete Information in Exchange Economies." *Econometrica* 57 (1): 115–34.

Palfrey, Thomas R., and Sanjay Srivastava. 1989b. "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem." *Journal of Political Economy* 97 (3): 668–91.

Postlewaite, Andrew, and David Schmeidler. 1986. "Implementation in Differential Information Economies." *Journal of Economic Theory* 39 (1): 14–33.

Repullo, R. 1987. "A Simple Proof of Maskin's Theorem on Nash Implementation." *Social Choice and Welfare* 4 (1): 39–41.

Saijo, Tatsuyoshi. 1988. "Strategy Space Reduction in Maskin's Theorem: Sufficient Conditions for Nash Implementation." *Econometrica* 56 (3): 693–700.

Wald, Abraham. 1945. "Statistical Decision Functions Which Minimize the Maximum Risk." *Annals of Mathematics* 46 (2): 265–80.

Wolitzky, Alexander. 2016. "Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade." *Theoretical Economics* 11 (3): 971–1004.