

# Mechanism Design With Ambiguous Communication Devices\*

Subir Bose<sup>†</sup> & Ludovic Renou<sup>‡</sup>

October 7, 2011

## Abstract

This paper considers mechanism design problems in environments with ambiguity-sensitive individuals. The novel idea is to introduce ambiguity in mechanisms so as to exploit the ambiguity sensitivity of individuals. We prove a revelation principle for the partial implementation of social choice functions by ambiguous mechanisms. We then revisit the classical monopolistic screening problem and show that ex-post full surplus extraction is possible, even when there is no ex-ante ambiguity.

**Keywords:** Mechanism design, ambiguity, revelation principle, communication device, multiple priors.

**JEL Classification Numbers:** C72; D71

---

\*We thank Rida Laraki, Vina Liu, Jérôme Renault and Tristan Tomala.

<sup>†</sup>Department of Economics, University of Leicester, Leicester LE1 7RH, United Kingdom.  
[sb345@le.ac.uk](mailto:sb345@le.ac.uk)

<sup>‡</sup>Department of Economics, University of Leicester, Leicester LE1 7RH, United Kingdom.  
[lr78@le.ac.uk](mailto:lr78@le.ac.uk)

# 1 Introduction

This paper studies mechanism design problems in environments with ambiguity-sensitive individuals. The central and novel aspect of our analysis is to consider *ambiguous* mechanisms. That is, we give the designer the possibility to *endogenously* engineer ambiguity in its mechanisms. After all, if individuals are ambiguity-sensitive, why should we preclude the use of ambiguous mechanisms, if it helps the designer in achieving its goals? As an example, consider a tax evasion problem. Tax authorities can be ambiguous about the likelihood of auditing taxpayers. For instance, suppose that the tax authority audits a taxpayer based on the draw of a ball from an urn containing a number of balls of different colors with unknown proportions (an “Ellsberg” urn). Would this dose of ambiguity help? As another example, consider the main refinancing operations of the European Central Bank (henceforth, ECB). To provide short-term liquidities, the ECB organizes weekly tender auctions, typically held every Tuesday. The ECB regulations carefully and meticulously specify all aspects of the tender auctions, so that there is little scope for ambiguity at this stage. However, prior to each weekly auction, the ECB communicates with the eligible counterparts about their liquidity needs.<sup>1</sup> And the communication can be ambiguous. Again, would such ambiguity help the ECB in steering the monetary policy? These questions echo the main theme of this paper: does the introduction of ambiguity in mechanism design help the designer in achieving its desired goals?

A natural starting point is to introduce ambiguity in the allocation rule. While simple and indeed appealing, the introduction of ambiguity in the allocation rule does not enlarge the set of (partially) implementable social choice functions, however.<sup>2,3</sup> The

---

<sup>1</sup>See “The implementation of Monetary Policy In The Euro Area,” February 2011, European Central Bank, <http://www.ecb.europa.eu/mopo/implement/omo/html/index.en.html>.

<sup>2</sup>In the sequel, we omit the qualifier “partial” and write “implementation” for “partial implementation.”

<sup>3</sup>See, however, the discussion towards the end of this section as to why the situation might be different if the goal is to implement a social choice *set*.

intuition is simple. While one can design an allocation rule that *formally* depends on some extraneous random variable (e.g., the draw from an “Ellsberg” urn), the (equilibrium) outcome has to be invariant across the realized value of this random variable, if the mechanism is to implement the given social choice function. Hence, a mechanism with an ambiguous allocation rule cannot (in itself) implement a social choice function that is not implementable by a classical (i.e., unambiguous) mechanism.

Nevertheless, we show that through the use of appropriately constructed ambiguous mechanisms, it is possible to implement social choice functions that are not implementable by using a standard, unambiguous mechanism. The critical ingredient in such construction is to add an *ambiguous* mediated communication stage prior to the *unambiguous* allocation stage. Classical mediated communication is comprised of sets of messages that agents can send to the designer, sets of messages that agents can receive from the designer, and a (i.e., *single*) mapping from messages received by the designer to probabilities over messages sent by the designer. Ambiguous mediated communication differs by having instead a *set* of mappings from messages received by the designer to probabilities over messages received. In essence, when an agent receives a message from the designer, he is ambiguous about the exact communication device that has been used by the designer. Our ambiguous mechanism is thus a mediated extension of a classical mechanism; however unlike a classical mediated extension (Myerson, 1986; Forges, 1986, 1990), the extensions we consider are ambiguous.

An ambiguous mechanism, i.e., a classical mechanism extended by rounds of ambiguous mediated communication, thus gives rise to a dynamic game with ambiguity-sensitive players. Our first task is to define an appropriate notion of equilibrium - called a “consistent planning equilibrium” - for such an environment. We next define the corresponding notion of incentive compatible social choice functions. The main theoretical contribution of this paper is to provide a suitable revelation principle for this class of mechanism design problems. An ambiguous “direct” mechanism is a direct revelation mechanism for the allocation stage, extended by one round of mediated ambiguous com-

munication. Our revelation principle (Theorem 1) states that a social choice function is implementable by an ambiguous mechanism if and only if it is implementable by an ambiguous “direct” mechanism.

This paper contributes to the growing literature on mechanism design with ambiguity-sensitive individuals, e.g., Bodoh-Creed (2011), Bose et al. (2006), Bose and Daripa (2009), Lopomo et al. (2010), Salo and Weber (1995), just to name a few. This literature differs from our analysis in one fundamental aspect, however. While we give the designer the possibility to endogenously engineer ambiguity, this literature constrains the designer to offer classical unambiguous mechanisms (often static, moreover). In particular, if there is *no* exogenous ambiguity, a social choice function is implementable by an unambiguous mechanism if and only if it is (classically) incentive compatible (Myerson, 1979). With no exogenous ambiguity and unambiguous mechanisms, it is as if individuals are ambiguity-insensitive. In contrast, even when there is no exogenous ambiguity, Proposition 1 demonstrates that the set of implementable social choice functions enlarges dramatically when ambiguous mechanisms are allowed. For instance, we revisit the classical monopolistic screening problem and show that ex-post full surplus extraction is possible, even when there is no ex-ante uncertainty.

After having completed this work, we became aware of a work in progress by Di Tillio, Kos and Messmer. These authors also explore the effects of introducing ambiguity in mechanisms in environments with ambiguity-sensitive preferences (more precisely, multiple-prior preferences). While the two works are closely related and indeed complement each others, there is a number of important differences. To start with, they focus on the revenue maximization problem in a single seller-buyer setup, whereas our objective is to characterize the entire set of social choice functions implementable by ambiguous mechanisms. Secondly, and more importantly, they introduce ambiguity in the allocation rule, so that the buyer is uncertain about the alternative to be implemented, even conditional on messages sent. As mentioned earlier, while we also consider the possibility of introducing ambiguity in the allocation rule, we show that it has no

effect for the implementation of social choice *functions* (see Section 6.2); this explains our focus on the introduction of ambiguity at the pre-allocation stage. Nonetheless, as their work suggests, introducing ambiguity in the allocation rule can expand the set of social choice *sets* that are implementable. Indeed, the revenue maximization problem of a seller, selling to a buyer privately informed about the valuation of the object, corresponds to the implementation of a set of social choice functions (and not to the implementation of a single social choice function). For instance, if the buyer has a continuum of possible valuations, any pair of social choice functions that differ in a set of measure zero are equivalent. Di Tillio et al. show that the introduction of ambiguity in the allocation rule makes it possible for the seller to extract more surplus from the trade, and when the buyer has a continuum of possible valuations, the seller can even extract the entire surplus (under the assumption that the distribution of valuations is atomless). Another important difference with our work is their focus on pure strategies. In contrast, we explicitly consider mixed strategies, which we think is an essential and fundamental component of any model with ambiguity averse agents.

We now present a simple example that illustrates the main arguments and intuitions of our analysis.

## 2 An Introductory Example

This section illustrates our main results with the help of a simple example. There are two players, labeled 1 and 2, two (payoff-relevant) types  $\theta$  and  $\theta'$  for each player, and four alternatives  $x$ ,  $y$ ,  $z$  and  $w$ . Types are private information. Throughout, we write “ $i$ ” for “ $i \in \{1, 2\}$ ” and omit the qualifier “ $j \neq i$ ,” whenever no confusion arises.

We assume that players have multiple-prior preferences (Gilboa and Schmeidler, 1989) and prior-by-prior updating (full Bayesian updating); see Pires (2002) for an axiomatization. Denote  $P_i(\theta)$  (resp.,  $P_i(\theta')$ ) the set of priors of player  $i$  of type  $\theta$  (resp.,  $\theta'$ ) and  $u_i$  his utility function. An element of  $P_i(\theta)$  (resp.,  $P_i(\theta')$ ) represents a prior belief of player  $i$  of type  $\theta$  (resp.,  $\theta'$ ) about the likelihood of player  $j$ 's type to be  $\theta$ . Utilities

are given in the table below. For instance, at state  $(\theta, \theta')$ , the utility of  $z$  to player 1 (resp., player 2) is 3 (resp., 0).

$(u_1, u_2)$	$(\theta, \theta)$	$(\theta, \theta')$	$(\theta', \theta)$	$(\theta', \theta')$
$x$	(2, 2)	(2, 0)	(0, 2)	(0, 0)
$y$	(1, 3)	(1, 2)	(0, 3)	(0, 2)
$z$	(3, 1)	(3, 0)	(2, 1)	(2, 0)
$w$	(0, 0)	(0, 2)	(2, 0)	(2, 2)

The designer aims at (partially) implementing the ex-post efficient social choice function  $f$  defined by:  $f(\theta, \theta) = x$ ,  $f(\theta, \theta') = y$ ,  $f(\theta', \theta) = z$  and  $f(\theta', \theta') = w$ .<sup>4</sup>

Assume that priors are independent of types, so that  $P_i(\theta) = P_i(\theta') := P_i$ . To crisply illustrate the role of ambiguity in mechanism design, we consider the starkest possible case, whereby there is no ex-ante ambiguity, i.e.,  $P_i$  is the singleton  $\{p_i\}$  for each player  $i$ . For concreteness, assume that  $\{p_i\} = \{2/3\}$ .

Clearly, no classical mechanism implements the social choice function  $f$ , since it fails to be incentive compatible. Yet, we argue that introducing some ambiguity in an otherwise classical mechanism makes it possible to implement  $f$ .

The central idea of this paper is to add a communication stage prior to the allocation stage and to introduce ambiguity at the communication stage.<sup>5</sup> More specifically, suppose that prior to the allocation stage, the players can communicate with the designer, i.e., they can send messages to and receive messages from the designer. A communication device specifies the messages players can send, the messages players can receive, and a probability system specifying the probabilities with which messages are sent to the players conditional on messages received from the players. A communication device is *ambiguous* if it specifies a (non-singleton) *set* of probability systems.

---

<sup>4</sup>Note that if we change the payoff of 3 to a payoff of  $3 - \varepsilon$  with  $\varepsilon > 0$  small enough, all our results go through and  $f$  is the unique ex-post efficient social choice function.

<sup>5</sup>As Section 6 demonstrates, ambiguity at the allocation stage does not help in implementing  $f$ .

To see how the ambiguous communication stage makes it possible to implement  $f$ , consider first the *hypothetical* situation where player  $i$  has ambiguous beliefs about the type of player  $j$ . Suppose that  $f$  is incentive compatible with respect to the set of beliefs  $\Pi_i^*$ , i.e.,

$$\begin{aligned} & \min_{\pi_i \in \Pi_i^*} u_i(f(\theta, \theta), \theta)\pi_i + u_i(f(\theta, \theta'), \theta)(1 - \pi_i) \geq \\ & \min_{\pi_i \in \Pi_i^*} [\sigma_i(u_i(f(\theta, \theta), \theta)\pi_i + u_i(f(\theta, \theta'), \theta)(1 - \pi_i)) + \\ & (1 - \sigma_i)(u_i(f(\theta', \theta), \theta)\pi_i + u_i(f(\theta', \theta'), \theta)(1 - \pi_i))], \end{aligned}$$

for all  $\sigma_i \in [0, 1]$ .<sup>6</sup> In words, truth-telling is optimal for player  $i$  of type  $\theta$  whenever player  $i$  expects player  $j$  to truthfully report his type and has the set of priors  $\Pi_i^*$  about the type of player  $j$ . For instance,  $f$  is incentive compatible with respect to  $\{0, 1\}$  or  $[0, 1]$ .<sup>7</sup> The crucial insight of this paper is that it is possible to construct an ambiguous communication device such that, conditional on every message player  $i$  can receive from the communication device, his posterior beliefs are precisely  $\Pi_i^*$ .

To illustrate further, we now construct an ambiguous communication device that generates the set of beliefs  $\Pi_i^* := \{0, 1\}$  for each player  $i$ . To this end, assume that each player can send  $\theta$  or  $\theta'$  to the designer, can receive  $\omega$  or  $\omega'$  from the designer and that there are two possible probability systems  $\lambda$  and  $\lambda'$ . Denote  $\lambda_i(\omega_i|\theta_j)$  the probability that player  $i$  receives the message  $\omega_i \in \{\omega, \omega'\}$  conditional on player  $j$ 's message  $\theta_j \in \{\theta, \theta'\}$  and let  $\lambda((\omega_1, \omega_2)|(\theta_1, \theta_2)) = \lambda_1(\omega_1|\theta_2)\lambda_2(\omega_2|\theta_1)$  for each possible  $(\omega_1, \omega_2)$  and  $(\theta_1, \theta_2)$ . Assuming  $\lambda_i(\omega|\theta) = 1$  and  $\lambda_i(\omega'|\theta') = 1$  thus fully defines the first probability system  $\lambda$ . Similarly, the second probability system  $\lambda'$  is fully specified by  $\lambda'_i(\omega|\theta') = 1$  and  $\lambda'_i(\omega'|\theta) = 1$ .<sup>8</sup> Clearly, if the probability system is  $\lambda$ , player  $i$ 's posterior belief is 1 if he receives the message  $\omega$  and 0, otherwise. Alternatively, if the probability system is

---

<sup>6</sup>Since it is strictly dominant for player  $i$  of type  $\theta'$  to truthfully report his type, we focus on player  $i$  of type  $\theta$ .

<sup>7</sup>More generally, it is incentive compatible for all  $[\underline{\pi}_i, \bar{\pi}_i]$  with  $\underline{\pi}_i \leq 1/2$ .

<sup>8</sup>Note that each probability system is a cyclic permutation of the other. This will play an important role in the general construction later.

$\lambda'$ , player  $i$ 's posterior belief is 0 if he receives the message  $\omega$  and 1, otherwise. Thus, regardless of the message received, player  $i$ ' set of posteriors is  $\Pi_i^* = \{0, 1\}$ .

It thus remains to argue that a stage of mediated and ambiguous communication prior to the allocation stage indeed implements  $f$ . Concretely, the implementing ambiguous mechanism is as follows. In the first stage, players communicate with the designer through the ambiguous communication device constructed above. In the second stage, players announce their types to the designer, who then implements an alternative according to  $f$ . It is important to note that the allocation depends only on the second-stage reports. The implementing mechanism is thus the classical direct mechanism extended by a phase of mediated communication.

So, assume that player  $j$  truthfully reports his type to the communication device at the first stage and truthfully reports his type at the second stage, regardless of the messages he has sent and received in the first stage. It follows from the construction of the communication device that player  $i$ 's beliefs at the second stage are given by  $\Pi_i^*$ , regardless of the messages he has sent and received in the first stage. Thus, player  $i$  has an incentive to truthfully reveal his type at the second stage. Lastly, since player  $i$  expects player  $j$  and himself to truthfully reveal their types at the allocation stage, regardless of the messages sent and received, he has no incentive to lie at the first stage. The social choice function  $f$  is thus implementable by an ambiguous mechanism.

While this construction might puzzle the reader, we now provide a decision-theoretic interpretation of our construction that shed light on important conceptual issues.

***A decision theoretic interpretation.*** Consider an urn containing 90 balls. Each ball is marked with either  $(\theta, \omega)$ ,  $(\theta, \omega')$ ,  $(\theta', \omega)$  or  $(\theta', \omega')$ . There are 60 balls marked with  $\theta$  and 30 balls marked with  $\theta'$ . Moreover, the composition of the urn is one of only two possible compositions. With the first composition, all balls marked with  $\theta$  (resp.,  $\theta'$ ) are also marked with  $\omega$  (resp.,  $\omega'$ ). With the second composition, all balls marked with  $\theta$  (resp.,  $\theta'$ ) are also marked with  $\omega'$  (resp.,  $\omega$ ).

A ball is drawn from the urn at random. The decision maker is offered two bets,  $A$



and  $B$ . The bet  $A$  gives  $x$  if the ball is marked with  $\theta$  and  $y$  if the ball is marked with  $\theta'$ , while the bet  $B$  gives  $z$  if the ball is marked with  $\theta$  and  $w$  if the ball is marked with  $\theta'$ . Prior to choosing a bet, the decision maker can observe whether the ball is marked with  $\omega$  or  $\omega'$ . The decision problem is represented in Figure 1; the first (resp., second) line corresponds to prizes in state  $\theta$  (resp.,  $\theta'$ ).

The decision maker is player 1 of type  $\theta$  in our mechanism design problem. The state space represents the possible types of player 2 and messages player 1 can receive from the communication device constructed above. Moreover, the possible composition of the urn respects the prior belief of player 1 as well as the ambiguity in the communication device. Finally, conditional on player 1 of type  $\theta$  expecting player 2 to tell the truth, the bet  $A$  corresponds to player 1 telling the truth, while the bet  $B$  corresponds to lying. Formally, the state space is  $\{\theta, \theta'\} \times \{\omega, \omega'\}$  and the set of prior beliefs of the decision maker is  $\{(2/3, 0, 0, 1/3), (0, 2/3, 1/3, 0)\}$ . We maintain the assumption of multiple prior preferences and assume prior-by-prior updating (full Bayesian updating).

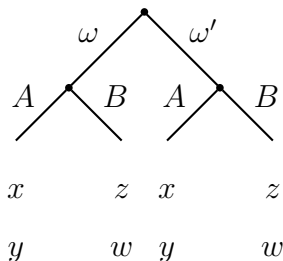


Figure 1: The decision problem

Clearly, upon learning whether the ball is marked with  $\omega$  or  $\omega'$ , the decision maker strictly prefers  $A$  over  $B$  and, more generally, over any randomization between  $A$  and  $B$ .

Consider now the ex-ante plans  $AA$ ,  $AB$ ,  $BA$  and  $BB$ , where the first (resp., second) letter corresponds to the choice of bet conditional on  $\omega$  (resp.,  $\omega'$ ). For instance, the plan  $AB$  prescribes the choice of  $A$  if  $\omega$  is revealed and the choice of  $B$  if  $\omega'$  is revealed. Assume that the decision maker evaluates the plan  $AB$  by “reducing” it to the bet giving

$x$  if the ball is  $(\theta, \omega)$ ,  $y$  if the ball is  $(\theta', \omega)$ ,  $z$  if the ball is  $(\theta, \omega')$ , and  $w$  if the ball is  $(\theta', \omega')$ . Similarly, for the other plans.

We have that the decision maker strictly prefers the plan  $BB$  to the plan  $AA$ , the plan  $AA$  to the plans  $AB$  and  $BA$ , and is indifferent between the plans  $AB$  and  $BA$ .<sup>9</sup>

To sum up, we have that conditional on either  $\omega$  or  $\omega'$ , the decision maker strictly prefers  $A$  to  $B$ , but ex-ante, he strictly prefers the plan  $BB$  to  $AA$ . The decision maker's preferences are *dynamically inconsistent* and our construction precisely exploits this fact.

We briefly comment on this fundamental aspect of our analysis and refer the reader to the special issue of *Economics and Philosophy* (2009) for an in-depth discussion and further references. Dynamic consistency and Bayesian updating are intimately related to Savage's sure-thing principle, and ambiguity-sensitive preferences generally entail a violation of the sure-thing principle. Consequently, if one wants to analyze ambiguity-sensitive preferences, then either dynamic consistency or full Bayesian updating must be relaxed, at least to some extent. The approach we follow in this paper is to relax the assumption of dynamic consistency. To analyze dynamic games with dynamically inconsistent preferences, we assume that players are *consistent planners*, i.e., at every information set a player is active, he chooses the best strategy given the strategies he will actually follow and the opponents' strategies (see Siniscalchi, 2010).

An alternative approach would be to maintain a form of dynamic consistency and to relax the assumption of full Bayesian updating. Hanany and Klibanoff (2007) provide such an alternative for the multiple-prior preferences. Without entering into details, their approach would require to update the prior  $(2/3, 0, 0, 1/3)$  upon observing  $\omega$  and the prior  $(0, 2/3, 1/3, 0)$  upon observing  $\omega'$ , so that the plan  $BB$  remains conditionally optimal.<sup>10</sup> Thus, according to their updating rule, the set of priors to be updated depend on the conditioning events (and, more generally, on the set of feasible plans and the unconditionally optimal plan considered). Whether one likes this feature or not, this is a logical implication of relaxing consequentialism so as to maintain dynamic consistency.

---

<sup>9</sup>More generally, the plan  $BB$  is preferred to any mixture over the plans  $AA$ ,  $AB$ ,  $BA$  and  $BB$ .

<sup>10</sup>We assume that the feasible set of plans is  $AA$ ,  $AB$ ,  $BA$  and  $BB$ .

We refer the reader to Siniscalchi (2009, 2010) and Al Najjar and Weinstein (2009) for more on this issue. Furthermore, we hasten to stress that a violation of dynamic consistency as defined in Hanany and Klibanoff (2007, axiom **DC**, p. 268) is *not* a necessary condition for our results to hold. To see this, let us modify the example so that  $u_1(w, \theta) = 1 = u_2(w, \theta)$ . With this modification, the decision maker strictly prefers  $BB$  to  $AA$  and, conditional on either  $\omega$  or  $\omega'$ , is indifferent between  $A$  and  $B$ . This does not violate axiom **DC** of Hanany and Klibanoff and yet  $f$  remains implementable by the ambiguous mechanism constructed above.

Yet, another alternative approach is to maintain consequentialism and (a form of) dynamic consistency, but to limit the possible attitude towards ambiguity. For instance, Epstein and Schneider (2001) provide a condition on the set of priors, called rectangularity, that guarantees the absence of preference reversals. In our example, their approach would require the set of priors to be

$$\{(2/3, 0, 0, 1/3), (0, 2/3, 1/3, 0), (1/3, 0, 0, 2/3), (0, 1/3, 2/3, 0)\}^{11}$$

Importantly to us, regardless of the strengths and weaknesses of those approaches, ambiguous mechanisms can implement social choice functions that are not ex-ante incentive compatible only if (a form of) dynamic inconsistency is assumed.

To conclude, we preview some secondary aspects of our analysis. Firstly, it is not difficult to see that our construction continues to apply when there is ex-ante ambiguity. For instance, suppose that player  $i$  set of priors is  $[\underline{p}_i, \bar{p}_i]$ . If  $\underline{p}_i > 1/2$ , the ambiguous communication device constructed above generates the required posteriors for the social choice function  $f$  to be incentive compatible (with respect to that set of posteriors). Alternatively, if  $\underline{p}_i \leq 1/2$ , the social choice function  $f$  is ex-ante incentive compatible and, thus, there is no need for ambiguous communication.

---

<sup>11</sup>With this set of priors, the social choice function  $f$  is implementable by a classical (unambiguous) direct mechanism. Alternatively, this follows from Epstein and Schneider's definition of dynamic consistency, which says that if  $A$  is conditionally preferred to  $B$ , conditional on both events  $\omega$  and  $\omega'$ , then  $AA$  is unconditionally preferred to  $BB$ .

Secondly, we have implicitly assumed that the decision maker considers only *two* possibilities: either the urn has the first composition or has the second composition. Alternatively, we might assume that the decision maker entertains *all* possibilities, so that his set of priors is the convex hull of  $(2/3, 0, 0, 1/3)$  and  $(0, 2/3, 1/3, 0)$ . This would not change our arguments: conditional on  $\omega$  or  $\omega'$ , the set of updated beliefs would include the posteriors  $(1, 0, 0, 0)$  and  $(0, 1, 0, 0)$ , and that is all we need for  $f$  to be implementable.

Lastly, our results extend naturally to a larger class of preferences. Section 6.1 elaborates on this issue.

### 3 Preliminaries

**Notations.** For any collection of  $n$  sets,  $Y_1, \dots, Y_n$ , we let  $Y := \times_{j \in \{1, \dots, n\}} Y_j$  and  $Y_{-i} := \times_{j \in \{1, \dots, n\} \setminus \{i\}} Y_j$ , with generic element  $y$  and  $y_{-i}$ , respectively. For any measurable space  $(Y, \mathcal{B}_Y)$ , we denote  $\Delta(Y)$  the set of probability measures over  $Y$ . Let  $\mu : Y \rightarrow \Delta(Y^*)$  be a measurable function between  $(Y, \mathcal{B}_Y)$  and  $(\Delta(Y^*), \mathcal{B}_{\Delta(Y^*)})$ . For any event  $E \in \mathcal{B}_{Y^*}$ , we write  $\mu(y)[E]$  for the probability of the event  $E$  according to  $\mu(y)$ . In the sequel, we assume that most sets (types, alternatives, messages, etc) are finite, so as to avoid unnecessary technicalities (e.g., measurability of strategies, conditioning on events with measure zero, etc). As section 6.4 demonstrates, our analysis extends to more general sets with appropriate measurability conditions.

**Environments.** An environment is a tuple  $\langle N, X, (\Theta_i)_{i \in N} \rangle$  where  $N := \{1, \dots, n\}$  is a finite set of  $n$  players,  $X$  a finite set of alternatives, and for each player  $i \in N$ ,  $\Theta_i$  is a finite set of payoff-relevant types. Types are privately known. A social choice function  $f : \Theta \rightarrow X$  assigns an alternative to each profile of types.

**Preferences and updating.** Let  $\mathcal{H}_i := \{h_i : \Theta_{-i} \rightarrow X\}$  be the set of player  $i$ 's acts. Player  $i$  of type  $\theta_i$  has type-dependent preferences over  $\Delta(\mathcal{H}_i)$ , the set of mixtures over acts; preferences are *ambiguity-sensitive*. Throughout most of the paper,

we assume that preferences have the maxmin expected utility representation (Gilboa and Schmeidler, 1989). More precisely, there exist a payoff function  $u_i : X \times \Theta \rightarrow \mathbb{R}$  and a non-empty, convex and closed valued correspondence  $P_i : \Theta_i \rightarrow \Delta(\Theta_{-i})$  such that player  $i$  of type  $\theta_i$  evaluates the act  $h_i$  as

$$\min_{p_i \in P_i(\theta_i)} \sum_{\theta_{-i} \in \Theta_{-i}} u_i(h_i(\theta_{-i}), \theta_i, \theta_{-i}) p_i(\theta_{-i}).$$

For simplicity, we assume full support, i.e.,  $P_i(\theta_i)$  is in the interior of  $\Delta(\Theta_{-i})$  for any  $\theta_i$ . Beliefs are type-independent if for each player  $i \in N$ , there exists  $P_i \in \Delta(\Theta_{-i})$  such that  $P_i(\theta_i) = P_i$  for all  $\theta_i \in \Theta_i$ . Type independence is weaker than stochastic independence.<sup>12</sup> We assume prior-by-prior updating (full Bayesian updating). Before going further, it is important to stress that our results do not crucially hinge on this particular representation of ambiguity-sensitive preferences. A number of other representations, such as  $\alpha$ -maxmin, minimax regret, variational preferences, deliver the same results. Section 6.1 critically discusses these assumptions and offers some generalizations.

**Ambiguous mechanisms.** Ambiguous mechanisms have two essential components: allocation mechanisms and ambiguous communication devices. An *allocation* mechanism is a pair  $\langle (M_i)_{i \in N}, g \rangle$  where for each player  $i$ ,  $M_i$  is a finite set of messages and  $g : \times_{i \in N} M_i \rightarrow X$  is an allocation rule. An allocation mechanism is thus a classical static mechanism.

A communication device is a tuple  $\langle (\Omega_i^*, \Omega_i)_{i \in N}, \lambda \rangle$ , where  $\Omega_i^*$  is a (finite) set of messages that player  $i$  can send to the communication device,  $\Omega_i$  a (finite) set of messages that player  $i$  can receive from the communication device and  $\lambda : \times_{i \in N} \Omega_i^* \rightarrow \Delta(\times_{i \in N} \Omega_i)$  a system of probability distributions:  $\lambda(\omega^*)[\omega]$  is the probability that the profile of messages  $\omega$  is sent to the players by the communication device, conditional on the profile of messages received  $\omega^*$ . An *ambiguous* communication device (for short, ambiguity device) is a tuple  $\langle (\Omega_i^*, \Omega_i)_{i \in N}, \Lambda \rangle$ , where  $\Lambda$  is a finite set of probability systems.

Given an allocation mechanism  $\langle M, g \rangle$ , we define the mediated extension of  $\langle M, g \rangle$

---

<sup>12</sup>Beliefs are stochastically independent if  $P_i = \times_{j \neq i} P_i^j$  with  $P_i^j \in \Delta(\Theta_j)$ .

as a mechanism in which  $T < \infty$  stages of mediated communication are allowed before  $\langle M, g \rangle$  is played. More precisely, there are  $T + 1$  stages. At stage  $t \in \{1, \dots, T\}$ , players communicate through the ambiguity device  $\langle (\Omega_{i,t}^*, \Omega_{i,t})_{i \in N}, \Lambda_t \rangle$ , i.e., each player  $i$  communicates a message  $\omega_{i,t}^* \in \Omega_{i,t}^*$  to the ambiguity device and receive a message  $\omega_{i,t} \in \Omega_{i,t}$ . Communication is private and simultaneous. At stage  $T + 1$ , players choose a message  $m_i \in M_i$  and the designer implements an alternative according to  $g$ . In words, prior to the allocation stage, players have the opportunity to communicate with each others through *mediated* communication devices. Communication is ambiguous and, moreover, does not directly influence the alternative implemented (cheap talk); the alternative implemented at the allocation stage  $T + 1$  depends only on the messages reported at stage  $T + 1$ . We call such a mechanism an *ambiguous mechanism*.

Fix an ambiguous mechanism  $\langle \langle (\Omega_{i,t}^*, \Omega_{i,t})_{i \in N}, \Lambda_t \rangle_{t=1, \dots, T}, \langle (M_i)_{i \in N}, g \rangle \rangle$ . Denote  $H_i^1 := \{\emptyset\}$  the initial history and for  $T+1 \geq t > 1$ ,  $H_i^t := \times_{\tau=1}^{t-1} (\Omega_{i,\tau}^* \times \Omega_{i,\tau})$  the set of all possible histories of messages sent and received up to stage  $t$  by player  $i$ . The set of terminal histories is  $\times_{i \in N} (H_i^{T+1} \times M_i)$ . Note that for any two terminal histories  $(h^{T+1}, m)$  and  $(\tilde{h}^{T+1}, \tilde{m})$  such that  $m = \tilde{m}$ , the alternative implemented is the same. However, expected payoffs might differ because of difference in beliefs.

**Behavioral strategies.** A behavioral strategy  $s_i$  for player  $i$  is a mapping  $s_i : (\cup_{t \leq T+1} H_i^t) \times \Theta_i \rightarrow (\cup_{t \leq T} \Delta(\Omega_{i,t}^*)) \cup \Delta(M_i)$  with  $s_i(h_i^t, \theta_i) \in \Delta(\Omega_{i,t}^*)$  for  $t \leq T$  and  $s_i(h_i^{T+1}, \theta_i) \in \Delta(M_i)$ .

**Assessments.** For each player  $i$ , for each stage  $t$ , for each history  $h_i^t$  and for each type  $\theta_i$ , we denote  $\Pi_i^{H, \Theta}(h_i^t, \theta_i) \subseteq \Delta(H_{-i}^t \times \Theta_{-i})$  the set of beliefs of player  $i$  of type  $\theta_i$ , conditional on the history  $h_i^t$ , about the types  $\theta_{-i}$  of his opponents and their private histories  $h_{-i}^t$ . The belief correspondence  $\Pi_i^{H, \Theta} : (\cup_{t \leq T+1} H_i^t) \times \Theta_i \rightarrow \cup_{t \leq T+1} \Delta(H_{-i}^t \times \Theta_{-i})$  with  $\Pi_i^{H, \Theta}(h_i^t, \theta_i) \subseteq \Delta(H_{-i}^t \times \Theta_{-i})$  for each  $(h_i^t, \theta_i)$  is called an assessment.

We impose the following two conditions on assessments. Firstly, assessments are consistent with initial priors, i.e.,  $\Pi_i^{H, \Theta}(\{\emptyset\}, \theta_i) = P_i(\theta_i) \times \mathbf{1}_{\{\emptyset\}}$  for each  $\theta_i \in \Theta_i$ , for each player  $i \in N$ . Secondly, assessments are consistent with “prior-by-prior” Bayesian

updating, whenever possible. More precisely, fix a profile of behavioral strategies  $(s_i)_i$ , a history  $h_i^t$  and a type  $\theta_i$  for player  $i$ . If  $(\omega_{i,t}^*, \omega_{i,t})$  has positive probability, i.e., there exist  $\pi_i(h_i^t, \theta_i) \in \Pi_i^{H,\Theta}(h_i^t, \theta_i)$  and  $\lambda_t \in \Lambda_t$  such that

$$\sum_{h_{-i}^t, \theta_{-i}, \omega_{-i,t}^*, \omega_{-i,t}} \pi_i(h_i^t, \theta_i)[(h_{-i}^t, \theta_{-i})] s_{-i}(h_{-i}^t, \theta_{-i})[\omega_{-i,t}^*] \lambda_t((\omega_{i,t}^*, \omega_{-i,t}^*))[(\omega_{i,t}, \omega_{-i,t})] > 0,$$

we update  $\pi_i(h_i^t, \theta_i)$  to  $\pi_i((h_i^t, (\omega_{i,t}^*, \omega_{i,t})), \theta_i)$ , with

$$\begin{aligned} & \pi_i((h_i^t, (\omega_{i,t}^*, \omega_{i,t})), \theta_i)[(h_{-i}^t, (\omega_{-i,t}^*, \omega_{-i,t})), \theta_{-i}] = \\ & \frac{\pi_i(h_i^t, \theta_i)[(h_{-i}^t, \theta_{-i})] s_{-i}(h_{-i}^t, \theta_{-i})[\omega_{-i,t}^*] \lambda_t((\omega_{i,t}^*, \omega_{-i,t}^*))[(\omega_{i,t}, \omega_{-i,t})]}{\sum_{h_{-i}^t, \theta_{-i}, \omega_{-i,t}^*, \omega_{-i,t}} \pi_i(h_i^t, \theta_i)[(h_{-i}^t, \theta_{-i})] s_{-i}(h_{-i}^t, \theta_{-i})[\omega_{-i,t}^*] \lambda_t((\omega_{i,t}^*, \omega_{-i,t}^*))[(\omega_{i,t}, \omega_{-i,t})]}. \end{aligned}$$

Thus, whenever  $(\omega_{i,t}^*, \omega_{i,t})$  has positive probability,  $\Pi_i^{H,\Theta}((h_i^t, (\omega_{i,t}^*, \omega_{i,t})), \theta_i)$  is obtained from  $\Pi_i^{H,\Theta}(h_i^t, \theta_i)$  by taking the union over all  $\pi_i(h_i^t, \theta_i) \in \Pi_i^{H,\Theta}(h_i^t, \theta_i)$  and  $\lambda_t \in \Lambda_t$  such that the above is well-defined. Alternatively, if  $(\omega_{i,t}^*, \omega_{i,t})$  has zero probability,  $\Pi_i^{H,\Theta}((h_i^t, (\omega_{i,t}^*, \omega_{i,t})), \theta_i)$  is unconstrained.

**Consistent planning equilibrium.** A consistent planning equilibrium (for short, equilibrium) is defined inductively by “backward induction.” It captures the idea that at stage  $t$ , a player chooses a plan of actions among the plan of actions that he will actually follow at later stages. Formally, a profile of strategies and assessments  $(s^*, \Pi^{H,\Theta})$  is an equilibrium if for each player  $i \in N$ , for each type  $\theta_i \in \Theta_i$  of player  $i$ , for all histories  $h_i^{T+1}$ ,  $s_i^*(h_i^{T+1}, \theta_i)$  satisfies the following:

$$\begin{aligned}
& \min_{\pi_i(h_i^{T+1}, \theta_i) \in \Pi_i^{H, \Theta}(h_i^{T+1}, \theta_i)} \left( \sum_{(\theta_{-i}, h_{-i}^{T+1})} \pi_i(h_i^{T+1}, \theta_i)[\theta_{-i}, h_{-i}^{T+1}] \right. \\
& \left. \left( \sum_{(m_i, m_{-i})} u_i(g(m_i, m_{-i}), \theta_i, \theta_{-i}) s_i^*(\theta_i, h_i^{T+1})[m_i] s_{-i}^*(\theta_{-i}, h_{-i}^{T+1})[m_{-i}] \right) \right) \geq \\
& \min_{\pi_i(h_i^{T+1}, \theta_i) \in \Pi_i^{H, \Theta}(h_i^{T+1}, \theta_i)} \left( \sum_{(\theta_{-i}, h_{-i}^{T+1})} \pi_i(h_i^{T+1}, \theta_i)[\theta_{-i}, h_{-i}^{T+1}] \right. \\
& \left. \left( \sum_{(m_i, m_{-i})} u_i(g(m_i, m_{-i}), \theta_i, \theta_{-i}) s_i^{T+1}[m_i] s_{-i}^*(\theta_{-i}, h_{-i}^{T+1})[m_{-i}] \right) \right),
\end{aligned}$$

for all  $s_i^{T+1} \in \Delta(M_i)$ . In words, the strategy prescribes the choice of a best response at each possible history  $h_i^{T+1}$ .

Denote  $U_i(s_i^*(\theta_i, h_i^{T+1}), s_{-i}^*(\theta_{-i}, h_{-i}^{T+1}), \theta_i, \theta_{-i})$  player  $i$ 's expected payoff under  $s^*$  when the history at stage  $T + 1$  is  $(h_i^{T+1}, h_{-i}^{T+1})$  and the type profile is  $(\theta_i, \theta_{-i})$ , i.e.,

$$\sum_{(m_i, m_{-i})} u_i(g(m_i, m_{-i}), \theta_i, \theta_{-i}) s_i^*(\theta_i, h_i^{T+1})[m_i] s_{-i}^*(\theta_{-i}, h_{-i}^{T+1})[m_{-i}].$$

We now move to stage  $T$ . Fix an history  $(h_i^T, h_{-i}^T)$  and a probability system  $\lambda_T \in \Lambda_T$ . Conditional on the type profile  $(\theta_i, \theta_{-i})$ , the subsequent history  $(h_i^{T+1}, h_{-i}^{T+1})$  is

$$((h_i^T, (\omega_{i,T}^*, \omega_{i,T})), (h_{-i}^T, (\omega_{-i,T}^*, \omega_{-i,T})))$$

with probability

$$s_i^*(h_i^T, \theta_i)[\omega_{i,T}^*] s_{-i}^*(h_{-i}^T, \theta_{-i})[\omega_{-i,T}^*] \lambda_T((\omega_{i,T}^*, \omega_{-i,T}^*))[(\omega_{i,T}, \omega_{-i,T})],$$

in which case the expected payoff is  $U_i(s_i^*(\theta_i, h_i^{T+1}), s_{-i}^*(\theta_{-i}, h_{-i}^{T+1}), \theta_i, \theta_{-i})$ . Thus, conditional on any history  $h_i^T$ , player  $i$ 's continuation payoff is well-defined and we can proceed with the definition of an equilibrium. At an equilibrium  $(s^*, \Pi^{H, \Theta})$ , for each player  $i \in N$ , for each type  $\theta_i \in \Theta_i$  of player  $i$ , for all histories  $h_i^T$ ,  $s_i^*(h_i^T, \theta_i)$  must satisfy the following:



$$\begin{aligned}
& \min_{(\lambda_T, \pi_i(h_i^T, \theta_i)) \in \Lambda_T \times \Pi_i^{H, \Theta}(h_i^T, \theta_i)} \sum_{(\theta_{-i}, h_{-i}^T)} \pi_i(h_i^T, \theta_i)[(\theta_{-i}, h_{-i}^T)] \\
& \left( \sum_{\Omega_T^* \times \Omega_T} (U_i(s_i^*(\theta_i, (h_i^T, (\omega_{i,T}^*, \omega_{i,T}))), s_{-i}^*(\theta_{-i}, (h_{-i}^T, (\omega_{-i,T}^*, \omega_{-i,T}))), \theta_i, \theta_{-i}) \right. \\
& \quad \left. s_i^*(h_i^T, \theta_i)[\omega_{i,T}^*] s_{-i}^*(h_{-i}^T, \theta_{-i})[\omega_{-i,T}^*] \lambda_T((\omega_{i,T}^*, \omega_{-i,T}^*))[(\omega_{i,T}, \omega_{-i,T})] \right) \geq \\
& \min_{(\lambda_T, \pi_i(h_i^T, \theta_i)) \in \Lambda_T \times \Pi_i^{H, \Theta}(h_i^T, \theta_i)} \sum_{(\theta_{-i}, h_{-i}^T)} \pi_i(h_i^T, \theta_i)[(\theta_{-i}, h_{-i}^T)] \\
& \left( \sum_{\Omega_T^* \times \Omega_T} (U_i(s_i^*(\theta_i, (h_i^T, (\omega_{i,T}^*, \omega_{i,T}))), s_{-i}^*(\theta_{-i}, (h_{-i}^T, (\omega_{-i,T}^*, \omega_{-i,T}))), \theta_i, \theta_{-i}) \right. \\
& \quad \left. s_i^T[\omega_{i,T}^*] s_{-i}^*(h_{-i}^T, \theta_{-i})[\omega_{-i,T}^*] \lambda_T((\omega_{i,T}^*, \omega_{-i,T}^*))[(\omega_{i,T}, \omega_{-i,T})] \right),
\end{aligned}$$

for all  $s_i^T \in \Delta(\Omega_{i,T}^*)$ . For each player  $i \in N$ , for each type  $\theta_i \in \Theta_i$ , for each history  $h_i^{T+1} \in H_i^{T+1}$  and  $h_i^T \in H_i^T$ , having defined  $s_i^*(h_i^{T+1}, \theta_i) \in \Delta(M_i)$  and  $s_i^*(h_i^T, \theta_i) \in \Delta(\Omega_{i,T})$ , we then proceed by “backward” induction to define the optimal strategy at each stage.

In a nutshell, a consistent planning equilibrium is similar to a perfect Bayesian equilibrium of a game between multiple selves of the same player with max-min expected utility. Notice that our specification of consistent planning differs slightly from the fully behavioral approach of Siniscalchi (2010). His specification incorporates a specific tie-breaking rule, stipulating that when a player has two or more optimal strategies at a given stage, ties are broken at earlier stages. Assuming such a tie-breaking rule would not substantially affect our results; only a minor modification to the concept of incentive compatibility (to be defined shortly) is needed.

A history  $h_i^{T+1}$  has positive probability under the profile of strategies  $s^*$  if there exist a sequence  $(\lambda_1, \dots, \lambda_T) \in \Lambda_1 \times \dots \times \Lambda_T$  and a type  $\theta_i \in \Theta_i$  such that  $h_i^{T+1}$  has strictly positive probability when evaluated according to  $(\lambda_1, \dots, \lambda_T)$  and  $(s_i(\theta_i, \cdot), s_{-i})$ . With these preliminaries done, we can now define the notion of implementation by ambiguous mechanisms.

**Definition 1** *The ambiguous mechanism  $\langle\langle(\Omega_{i,t}^*, \Omega_{i,t})_{i \in N}, \Lambda_t\rangle_{t=1,\dots,T}, \langle(M_i)_{i \in N}, g\rangle\rangle$  (partially) implements the social choice function  $f$  if there exists a consistent planning equilibrium  $(s^*, \Pi^{H,\Theta})$  such that*

$$g(m_i, m_{-i}) = f(\theta_i, \theta_{-i}),$$

*for all  $(m_i, m_{-i})$  with  $\times_{i \in N} s_i^*(\theta_i, h_i^{T+1})[m_i] > 0$ , for all  $(h_i^{T+1}, h_{-i}^{T+1})$  having positive probability under  $s^*$ .*

According to Definition 1, for a social choice function to be implementable, there must exist an ambiguous mechanism and an equilibrium such that for all histories of messages sent to and received from the communication devices, for all messages sent at the allocation stage following these histories, the designer implements the correct outcome. The aim of this paper is to characterize the set of social choice functions that are implementable by ambiguous mechanisms.

## 4 Revelation Principle

This section presents our main results. We first define incentive compatibility.

### 4.1 Incentive Compatibility

A social choice function is incentive compatible if players have an incentive not only to truthfully reveal their types at the allocation stage when they expect others to do so, but also to generate the information (and, thus, beliefs) designed for their types at the communication stage.

**Definition 2** *A social choice function  $f$  is incentive compatible for player  $i \in N$  if there exists a non-empty valued correspondence  $\Pi_i : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Theta_{-i})$  such that for each  $\theta_i \in \Theta_i$ ,*

$$\begin{aligned}
& \min_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) \geq \\
& \min_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{\theta'_i \in \Theta_i} \sigma_i(\theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}),
\end{aligned} \tag{IC}$$

for all  $\sigma_i(\theta_i) \in \Delta(\Theta_i)$ , for all  $\theta_i'' \in \Theta_i$ . The social choice function  $f$  is  $\Pi_i$ -incentive compatible if it is incentive compatible for player  $i$  with respect to the belief correspondence  $\Pi_i$ . The social choice function  $f$  is incentive compatible if it is incentive compatible for each player  $i \in N$ .

Intuitively,  $\Pi_i(\theta_i'', \theta_i)$  is the set of beliefs player  $i$  of type  $\theta_i$  has at the allocation stage when he communicates as type  $\theta_i''$  at the communication stage. Incentive compatibility thus requires truth-telling to be a “max-min” equilibrium of the direct allocation mechanism, even when players are untruthful at the communication stage. Consequently, players have an incentive to be truthful at the communication stage. Two additional remarks are worth making. Firstly, Definition 2 explicitly considers mixed strategies. This is crucial as ambiguity-sensitive preferences frequently display ambiguity aversion, for which hedging is essential. Secondly, with a slight abuse of notation, let  $P_i(\cdot, \theta_i) \equiv P_i(\theta_i)$  for all  $\theta_i$ . Clearly, if the social choice function  $f$  is  $P_i$ -incentive compatible for each player  $i$ , then it is implementable by an ambiguous mechanism. Yet, as the example in section 2 shows, incentive compatibility with respect to the prior beliefs is not necessary for the implementation of social choice functions by finite mechanisms. The rest of this section is devoted to the full characterization of implementable social choice functions by finite ambiguous mechanisms.

## 4.2 The Main Theorem

For any ambiguity device  $\langle (\Theta_i, \Omega_i)_{i \in N}, \Lambda \rangle$ , define the belief function  $\zeta_{\theta_i} : \Theta_i \times P_i(\theta_i) \times \Omega_i \times \Lambda \rightarrow \Delta(\Theta_{-i})$  as

$$\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda)[\theta_{-i}] = \frac{\sum_{\omega_{-i}} \lambda((\theta'_i, \theta_{-i}))[(\omega_i, \omega_{-i})] p_{\theta_i}(\theta_{-i})}{\sum_{\omega_{-i}} \sum_{\theta_{-i}} \lambda((\theta'_i, \theta_{-i}))[(\omega_i, \omega_{-i})] p_{\theta_i}(\theta_{-i})}.$$

So,  $\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda) \in \Delta(\Theta_{-i})$  is the posterior belief of player  $i$  of type  $\theta_i$  when he sends the message  $\theta'_i$  to and receives the message  $\omega_i$  from the communication device  $\langle (\Theta_i, \Omega_i)_{i \in N}, \lambda \rangle$ , and has prior beliefs  $p_{\theta_i}$  (conditional on player  $i$ 's opponents telling the truth). Similarly, for any  $\lambda_i : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$ , we denote  $\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i) \in \Delta(\Theta_{-i})$  the posterior belief of player  $i$  of type  $\theta_i$  when he sends the message  $\theta'_i$  to and receives the message  $\omega_i$  from the communication device  $\langle \Theta_i, \Omega_i, \lambda_i \rangle$ , and has prior beliefs  $p_{\theta_i}$ .

**Theorem 1** *The following statements are equivalent:*

1. *The social choice function  $f$  is implementable by a finite ambiguous mechanism.*
2. *For each player  $i \in N$ , there exists a finite family of non-empty valued correspondences  $(\Pi_i^k : \Theta_i \times \Theta_i \rightarrow \Delta(\Theta_{-i}))_{k \in K_i}$  such that:*

(IC) *For each  $k \in K_i$ , the social choice function  $f$  is  $\Pi_i^k$ -incentive compatible.*

(B) *There exist  $\Omega_i := \cup_{k \in K_i} \Omega_i^k$  and  $\lambda_i : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$  such that*

$$\cup_{p_{\theta_i} \in P(\theta_i)} \cup_{\omega_i \in \Omega_i^k} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i)\} = \Pi_i^k(\theta'_i, \theta_i), \quad (\text{B})$$

*for all  $\theta'_i \in \Theta_i$ , for all  $\theta_i \in \Theta_i$ , for all  $k \in K_i$ .*

Before presenting the formal proof, let us comment on Theorem 1. Theorem 1 is a revelation principle; it states necessary and sufficient conditions for a social choice function to be implementable by a finite ambiguous mechanism. More specifically, if a social choice function is implementable, then it is incentive compatible with respect to some sets of beliefs (condition (IC)). Moreover, those sets of beliefs must result from the mediated communication and, thus, must satisfy a martingale property (condition (B)). An important element of (the necessity part of) the proof is the observation that if a social choice function  $f$  is implementable with  $T$  stages of mediated communication, then it is implementable with a single stage of mediated communication. While intuitive

and indeed well-known in the literature on mediated communication, this observation is not as trivial as it may appear in a world with dynamically inconsistent preferences. To get some intuitions, consider the following decision problem. There is a single player, two payoff-relevant types  $\theta$  and  $\theta'$ , two messages  $\omega$  and  $\omega'$ , and two acts  $f_1$  and  $f_2$  (acts map  $\Theta_{-i}$  to  $X$ ), with  $f_1 \succ_{\{\omega\}}^\theta f_2$ ,  $f_1 \succ_{\{\omega'\}}^\theta f_2$  at state  $\theta$ , and  $f_2 \succ_{\{\omega\}}^{\theta'} f_1$ ,  $f_2 \succ_{\{\omega'\}}^{\theta'} f_1$  at state  $\theta'$  ( $\succ_h^\theta$  (resp.,  $\succ_h^{\theta'}$ ) denotes the (resp., strict) preference relation at state  $\theta$  conditional on history  $h$ ). We want to “implement”  $f_1$  at state  $\theta$  and  $f_2$  at state  $\theta'$ . See the left panel of Figure 2 for the decision problem.

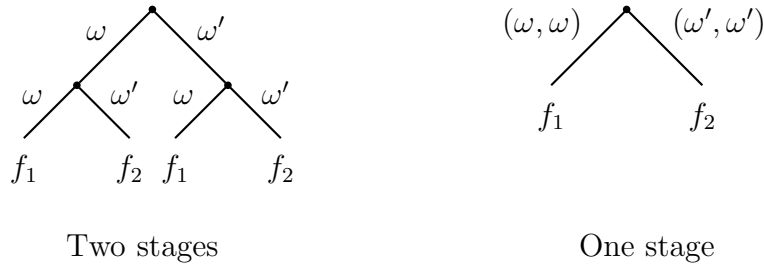


Figure 2: The decision problem

Clearly, it is a consistent planning equilibrium for the decision maker of type  $\theta$  (resp.,  $\theta'$ ) to announce  $\omega$  (resp.,  $\omega'$ ) at all histories, thus implementing  $f_1$  at state  $\theta$  and  $f_2$  at state  $\theta'$ . Alternatively, consider the one-period decision problem on the right panel of Figure 2. With dynamically consistent preferences, we have  $f_1 \succ_{\{\emptyset\}}^\theta f_2$  and  $f_2 \succ_{\{\emptyset\}}^{\theta'} f_1$ , so that it is an equilibrium for the decision maker of type  $\theta$  (resp.,  $\theta'$ ) to announce  $(\omega, \omega)$  (resp.,  $(\omega', \omega')$ ), thus also implementing  $f_1$  at state  $\theta$  and  $f_2$  at state  $\theta'$ . However, if preferences are dynamically inconsistent, we might have  $f_2 \succ_{\{\emptyset\}}^\theta f_1$  or  $f_1 \succ_{\{\emptyset\}}^{\theta'} f_2$ , so that the one-period decision problem is not “equivalent” to the two-period decision problem. This issue, however, does not create difficulties in our framework. Recall that our definition of implementation requires the correct outcomes to be implemented at all histories of messages with positive probabilities. As applied to this decision problem, our definition thus requires the correct outcome to be implemented at  $(\omega, \omega)$  and  $(\omega', \omega')$ , which essentially requires  $f_1$  to be equal to  $f_2$ .

Conversely, to implement a social choice function that satisfies conditions (IC) and (B), we construct a simple two-stage mechanism that consists of an allocation stage and a stage of mediated communication prior to the allocation stage. The communication stage generates the appropriate posterior beliefs at the allocation stage, so as to guarantee truth-telling. To be more precise, suppose that the social choice function  $f$  is incentive compatible with respect to  $\Pi_i$ , with  $\Pi_i$  satisfying condition (B). Thus, there exists a communication device  $\langle \Theta_i, \Omega_i, \lambda_i^* \rangle$  such that the set of *all* posteriors of player  $i$  of type  $\theta_i$  is  $\Pi_i(\theta'_i, \theta_i)$  whenever he sends  $\theta'_i$  to the device and others truthfully report their types. Yet, this does not imply that the set of posteriors is  $\Pi_i(\theta'_i, \theta_i)$  conditional on any  $\omega_i$  and, consequently, this does not guarantee truth-telling at the allocation stage which is guaranteed only if the set of posteriors conditional on any  $\omega_i$  is  $\Pi_i(\theta'_i, \theta_i)$ . Consider a permutation  $\rho : \Omega_i \rightarrow \Omega_i$  and define  $\lambda_i^\rho$  by  $\lambda_i^\rho(\theta_i, \theta_{-i})[\omega_i] = \lambda_i^*(\theta_i, \theta_{-i})[\rho(\omega_i)]$ . We have that  $\cup_{p_{\theta_i} \in P(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i^\rho)\} = \cup_{p_{\theta_i} \in P(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \rho(\omega_i), \lambda_i^*)\}$ . The idea is then to consider the set of communication devices  $\Lambda_i$  as the union over all cyclic permutations of  $\{\lambda_i^\rho\}$ ; this guarantees that the set of posteriors conditional on any  $\omega_i$  is  $\Pi_i(\theta'_i, \theta_i)$ , as required.

The classical revelation principle may be viewed as a corollary of Theorem 1. To see this, note that if we restrict the beliefs' correspondences to be single-valued, the model is indistinguishable from subjective expected utility. So, let us assume that  $P_i(\theta_i) = \{p_{\theta_i}\}$  for all players  $i \in N$ , for all types  $\theta_i \in \Theta_i$ , and suppose that the second statement in Theorem 1 is true with each  $\Pi_i^k$  being single-valued. The  $\Pi_i^k$ -incentive compatibility of  $f$  implies that for all  $\theta_i \in \Theta_i$ , for all  $\theta'_i \in \Theta_i$ , for all  $\theta''_i \in \Theta_i$ ,

$$\sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_{\theta''_i, \theta_i}^k[\theta_{-i}] \geq \sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_{\theta''_i, \theta_i}^k[\theta_{-i}],$$

with  $\{\pi_{\theta''_i, \theta_i}^k\} := \Pi_i^k(\theta''_i, \theta_i)$ . The classical incentive compatibility of  $f$  then follows from

condition (B), i.e., for all  $i \in N$ , for all  $\theta_i \in \Theta_i$ , for all  $\theta'_i \in \Theta_i$ ,

$$\sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) p_{\theta_i}[\theta_{-i}] \geq \sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) p_{\theta_i}[\theta_{-i}].^{13}$$

Conversely, if the social choice function  $f$  is classically incentive compatible, then the second statement of Theorem 1 trivially holds with  $\Pi_i(\cdot, \theta_i) \equiv \{p_{\theta_i}\}$  for each  $\theta_i$ .

**Proof** (2)  $\Rightarrow$  (1). The proof is constructive. Suppose that for each player  $i \in N$ , there exists a unique non-empty valued correspondence  $\Pi_i : \Theta_i \times \Theta_i \rightarrow \Delta(\Theta_{-i})$  such that conditions (IC) and (B) hold. (We treat the general case of a family of correspondences in Appendix.)

We consider a two-stage ambiguous mechanism, where players announce a type to the communication device in the first stage, receive messages in the first stage conditional on their announcements, and again report a type in the second stage. Formally, we consider the ambiguity device  $\langle \times_{i \in N} \Theta_i, \times_{i \in N} \Omega_i, \Lambda \rangle$  and the (direct) allocation mechanism  $\langle \times_{i \in N} \Theta_i, f \rangle$ . The first step in the proof consists in constructing  $\Lambda$ .

**Step 1: Construction of  $\Lambda$ .** Since condition (B) holds, for each player  $i \in N$ , there exist  $\Omega_i$  and  $\lambda_i^* : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$  such that

$$\cup_{\omega_i \in \Omega_i} \cup_{p_{\theta_i} \in P(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i^*)\} = \Pi_i(\theta'_i, \theta_i),$$

for all  $\theta'_i \in \Theta_i$ , for all  $\theta_i \in \Theta_i$ . Consider a permutation  $\rho : \Omega_i \rightarrow \Omega_i$  and define  $\lambda_i^\rho : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$  with  $\lambda_i^\rho((\theta_i, \theta_{-i}))[\omega_i] = \lambda_i^*((\theta_i, \theta_{-i}))[\rho(\omega_i)]$  for all  $\omega_i$ , for all  $\theta_{-i}$ . Let  $\Lambda_i$  be the collection of all  $\lambda_i^\rho$  such that  $\rho$  is a cyclic permutation (the cardinality of  $\Lambda_i$  is thus the cardinality of  $\Omega_i$ ). It follows that

$$\cup_{\lambda_i \in \Lambda_i} \cup_{p_{\theta_i} \in P(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i^\rho)\} = \cup_{\omega_i \in \Omega_i} \cup_{p_{\theta_i} \in P(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i^*)\} = \Pi_i(\theta'_i, \theta_i),$$

<sup>13</sup>More precisely, for each  $\theta_i \in \Theta_i$ , for each  $\theta''_i \in \Theta_i$ , we have

$$\underbrace{\sum_{k \in K_i} \left( \sum_{\omega_i \in \Omega_i^k} \sum_{\omega_{-i} \in \Omega_{-i}} \sum_{\theta_{-i} \in \Theta_{-i}} \lambda_i(\theta''_i, \theta_{-i})[(\omega_i, \omega_{-i})] p_{\theta_i}[\theta_{-i}] \pi_{\theta''_i, \theta_i}^k \right)}_{\lambda_{\theta''_i, \theta_i}^k} = p_{\theta_i},$$

with  $\sum_{k \in K_i} \lambda_{\theta''_i, \theta_i}^k = 1$  and  $\lambda_{\theta''_i, \theta_i}^k \geq 0$  for each  $k \in K_i$ .

for all  $\omega_i \in \Omega_i$ , for all  $\theta'_i \in \Theta_i$ , for all  $\theta_i \in \Theta_i$ . Lastly, for each  $i \in N$ , choose  $\lambda_i \in \Lambda_i$  and define  $\lambda : \times_{i \in N} \Theta_i \rightarrow \Delta(\times_{i \in N} \Omega_i)$  with  $\lambda((\theta_i)_{i \in N})[(\omega_i)_{i \in N}] = \times_{i \in N} \lambda_i((\theta_i, \theta_{-i}))[\omega_i]$ . The set  $\Lambda$  is the collection of all such  $\lambda$ .

**Step 2: Construction of the equilibrium.** Consider the following profile of strategies. At the initial history, players truthfully report their types to the ambiguity device, i.e.,  $s_i^*(\{\emptyset\}, \theta_i) = \theta_i$  for all  $\theta_i \in \Theta_i$ , and for any history  $(\theta'_i, \omega_i)$  of messages sent and received at the communication stage, player  $i$  of type  $\theta_i$  truthfully reports his type at the allocation stage, i.e.,  $s_i^*((\theta'_i, \omega_i), \theta_i) = \theta_i$  for all histories  $(\theta'_i, \omega_i) \in \Theta_i \times \Omega_i$  of messages sent and received and type  $\theta_i \in \Theta_i$ . Assessments are consistent with Bayes rule, whenever possible. More precisely, suppose that the history  $(\theta'_i, \omega_i)$  has positive probability, i.e., there exist  $\lambda_i$  and  $\theta_{-i}$  such that  $\lambda_i(\theta'_i, \theta_{-i})[\omega_i] > 0$ . Define the posterior  $\pi_i^{H, \Theta}((\theta'_i, \omega_i), \theta_i)$  conditional on the history  $(\omega_i, \theta'_i)$  and type  $\theta_i$  by

$$\pi_i^{H, \Theta}((\theta'_i, \omega_i), \theta_i)[(\omega_{-i}, \theta_{-i})] = \frac{\lambda_i(\theta'_i, \theta_{-i})[\omega_i] \lambda_{-i}(\theta'_i, \theta_{-i})[\omega_{-i}] p_{\theta_i}[\theta_{-i}]}{\sum_{\theta_{-i}, \omega_{-i}} \lambda_i(\theta'_i, \theta_{-i})[\omega_i] \lambda_{-i}(\theta'_i, \theta_{-i})[\omega_{-i}] p_{\theta_i}[\theta_{-i}]},$$

and let  $\Pi_i^{H, \Theta}((\theta'_i, \omega_i), \theta_i)$  be the union over all  $\lambda_i \in \Lambda_i$  for which there exists  $\theta_{-i}$  such that  $\lambda_i(\theta'_i, \theta_{-i})[\omega_i] > 0$ . If the history  $(\theta'_i, \omega_i)$  has zero probability, simply let  $\Pi_i^{H, \Theta}((\theta'_i, \omega_i), \theta_i)$  to be equal to  $\Pi_i^{H, \Theta}((\theta'_i, \omega_i^*), \theta_i)$  for some  $(\theta'_i, \omega_i^*)$  with positive probability (at least one exists).

**Step 3: No profitable deviation.** By construction of the ambiguity device, we have that the beliefs of player  $i$  of type  $\theta_i$  about the types  $\theta_{-i}$  of his opponents are  $\Pi_i(\theta'_i, \theta_i)$  at history  $(\theta'_i, \omega_i)$ , regardless of  $\omega_i$ . Since the social choice function is  $\Pi_i$ -incentive compatible for each player  $i \in N$ , it follows that no player has a profitable deviation from truth-telling at any history  $(\theta'_i, \omega_i) \in \Theta_i \times \Omega_i$ . Finally, no player has a profitable deviation at the first stage, since they expect  $f$  to be implemented in the second stage, regardless of their first stage announcement. This completes this part of the proof.

(1)  $\Rightarrow$  (2). Suppose that the social choice function  $f$  is implementable by the am-



biguous mechanism

$$\langle\langle (\hat{\Omega}_{i,t}^*, \hat{\Omega}_{i,t})_{i \in N}, \hat{\Lambda}_t \rangle_{t=1, \dots, T}, \langle\langle (M_i)_{i \in N}, g \rangle\rangle\rangle.$$

From Lemma 1 in Appendix,  $f$  is implementable by a two-stage mechanism, in which players directly and truthfully report their types to the communication device at the communication stage. Denote the two-stage mechanism by

$$\langle\langle (\Theta_i, \Omega_i)_{i \in N}, \Lambda \rangle, \langle\langle (M_i)_{i \in N}, g \rangle\rangle\rangle.$$

Thus, there exists a consistent planning equilibrium  $(s^*, \Pi^{H, \Theta})$  such that  $g(m_i, m_{-i}) = f(\theta_i, \theta_{-i})$  for all  $(m_i, m_{-i})$  with  $s_i^*(h_i, \theta_i)[m_i]s_{-i}^*(h_{-i}, \theta_{-i})[m_{-i}] > 0$ , for all  $(h_i, h_{-i})$  having positive probability under  $s^*$ , and  $s_i^*(\{\emptyset\}, \theta_i)[\theta_i] = 1$  for all  $\theta_i$ . Observe that a non-terminal history  $h_i \neq \{\emptyset\}$  for player  $i$  is a message sent and received, i.e.,  $(\theta'_i, \omega'_i)$ . Moreover, the history  $(\theta'_i, \omega'_i)$  has positive probability if there exist  $\lambda \in \Lambda$  and  $\theta'_{-i} \in \Theta_{-i}$  such that  $\sum_{\omega_{-i}} \lambda(\theta'_i, \theta'_{-i})[(\omega'_i, \omega_{-i})] > 0$ .<sup>14</sup>

Consider player  $i$  of type  $\theta_i$  and any history  $h_i = (\theta_i, \omega_i)$  having positive probability. By definition of a consistent planning equilibrium, we have that

$$\begin{aligned} & \min_{\pi_i(h_i, \theta_i) \in \Pi_i^{H, \Theta}(h_i, \theta_i)} \left( \sum_{(\theta_{-i}, h_{-i})} \pi_i(h_i, \theta_i)[\theta_{-i}, h_{-i}] \right. \\ & \left. \left( \sum_{(m_i, m_{-i})} u_i(g(m_i, m_{-i}), \theta_i, \theta_{-i}) s_i^*(h_i, \theta_i)[m_i] s_{-i}^*(h_{-i}, \theta_{-i})[m_{-i}] \right) \right) \geq \\ & \min_{\pi_i(h_i, \theta_i) \in \Pi_i^{H, \Theta}(h_i, \theta_i)} \left( \sum_{(\theta_{-i}, h_{-i})} \pi_i(h_i, \theta_i)[\theta_{-i}, h_{-i}] \right. \\ & \left. \left( \sum_{(m_i, m_{-i})} u_i(g(m_i, m_{-i}), \theta_i, \theta_{-i}) s_i[m_i] s_{-i}^*(h_{-i}, \theta_{-i})[m_{-i}] \right) \right), \end{aligned}$$

for all  $s_i \in \Delta(M_i)$ . In particular, this is true for any deviation  $s_i$  such that  $s_i(\theta_i, h_i) = \sum_{\theta'_i} \sigma_i(\theta_i)[\theta'_i] s_i^*(\theta'_i, h_i)$  with  $\sigma_i(\theta_i) \in \Delta(\Theta_i)$  and  $s_i$  coincides with  $s_i^*$ , otherwise. Thus, we

---

<sup>14</sup>Because of the assumption of full support, this is equivalent to: there exist  $\lambda \in \Lambda$  and  $p_{\theta_i} \in P_i(\theta_i)$  such that  $\sum_{\theta_{-i}} \sum_{\omega_{-i}} \lambda(\theta'_i, \theta'_{-i})[(\omega'_i, \omega_{-i})] s_{-i}^*(\theta_{-i})[\theta'_{-i}] p_{\theta_i}(\theta_{-i}) > 0$ .

have

$$\min_{\pi_i(h_i, \theta_i) \in \Pi_i^{H, \Theta}(h_i, \theta_i)} \sum_{(\theta_{-i}, h_{-i})} \pi_i(h_i, \theta_i) [\theta_{-i}, h_{-i}] u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \geq$$

$$\min_{\pi_i(h_i, \theta_i) \in \Pi_i^{H, \Theta}(h_i, \theta_i)} \sum_{(\theta_{-i}, h_{-i})} \pi_i(h_i, \theta_i) [\theta_{-i}, h_{-i}] \left( \sum_{\theta'_i} \sigma_i(\theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \right),$$

for all  $\sigma_i(\theta_i) \in \Delta(\Theta_i)$ . It follows that there exists a set  $\Pi_i^k(\theta_i, \theta_i) \subseteq \Delta(\Theta_{-i})$  of beliefs of player  $i$  of type  $\theta_i$  such that

$$\min_{\pi_i \in \Pi_i^k(\theta_i, \theta_i)} \sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) \geq$$

$$\min_{\pi_i \in \Pi_i^k(\theta_i, \theta_i)} \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{\theta'_i \in \Theta_i} \sigma_i(\theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}),$$

for all  $\sigma_i(\theta_i) \in \Delta(\Theta_i)$ . More precisely, for any history  $h_i$  and type  $\theta_i$ , denote  $\Pi_i^\Theta(h_i, \theta_i) := \cup_{\pi_i(h_i, \theta_i) \in \Pi_i^{H, \Theta}(h_i, \theta_i)} \{\sum_{h_{-i}} \pi_i(h_i, \theta_i) [(h_{-i}, \theta_{-i})]\}$  and let  $\Pi_i^k(\theta_i, \theta_i) = \Pi_i^\Theta((\theta_i, \omega_i), \theta_i)$ . With any other history  $(\theta_i, \omega'_i)$  such that  $\Pi_i^\Theta((\theta_i, \omega'_i), \theta_i) = \Pi_i^\Theta((\theta_i, \omega_i), \theta_i)$ , we thus associate the same set of beliefs  $\Pi_i^k(\theta_i, \theta_i)$ . Note that  $\Pi_i^\Theta(h_i, \theta_i) = \cup_{\lambda \in \Lambda} \cup_{p_{\theta_i} \in P_i(\theta_i)} \{\zeta_{\theta_i}(\theta_i, p_{\theta_i}, \omega_i, \lambda)\}$ , when  $h_i = (\theta_i, \omega_i)$ .

Similarly, consider any history  $h_i = (\theta'_i, \omega_i)$  having positive probability. As above, there exists a set of beliefs of player  $i$  of type  $\theta_i$ ,  $\Pi_i^k(\theta'_i, \theta_i)$ , such that

$$\min_{\pi_i \in \Pi_i^k(\theta'_i, \theta_i)} \sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) \geq$$

$$\min_{\pi_i \in \Pi_i^k(\theta'_i, \theta_i)} \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{\theta'_i \in \Theta_i} \sigma_i(\theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}),$$

for all  $\sigma_i(\theta_i) \in \Delta(\Theta_i)$ . Replicating the arguments for each type of player  $i$  and each history thus implies the existence of a finite family of non-empty valued correspondences  $(\Pi_i^k : \Theta_i \times \Theta_i \rightarrow \Delta(\Theta_{-i}))_{k \in K_i}$  such that the social choice function  $f$  is  $\Pi_i^k$ -incentive compatible, for each  $k \in K_i$ .

To complete the proof, for each  $\lambda \in \Lambda$ , define  $\lambda_i : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$  with for all  $(\theta_i, \theta_{-i})$ ,

$$\lambda_i(\theta_i, \theta_{-i})[\omega_i] := \sum_{\omega_{-i} \in \Omega_{-i}} \lambda(\theta_i, \theta_{-i})[(\omega_i, \omega_{-i})],$$

and let  $\Lambda_i$  be the set of all such  $\lambda_i$ . Let  $\Omega_i^* := \Omega_i \times \Lambda_i$  and define  $\lambda_i^* : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i^*)$  as follows:

$$\lambda_i^*(\theta_i, \theta_{-i})[(\omega_i, \lambda_i)] := \frac{1}{|\Lambda_i|} \lambda_i(\theta_i, \theta_{-i})[\omega_i],$$

for all  $(\theta_i, \theta_{-i})$ . Lastly, for each  $k$ , define  $\Omega_i^k$  as  $\{\omega_i : \Pi_i^\Theta((\theta'_i, \omega_i), \theta_i) = \Pi_i^k(\theta'_i, \theta_i)\}$  and let  $\Omega_i^{*,k} = \Omega_i^k \times \Lambda_i$ . Note that  $\cup_k \Omega_i^k = \Omega_i$  and thus  $\cup_k \Omega_i^{*,k} = \Omega_i^*$ . It follows that

$$\begin{aligned} \cup_{\omega_i^* \in \Omega_i^{*,k}} \cup_{p_{\theta_i} \in P_i(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i^*, \lambda_i^*)\} &= \cup_{\omega_i \in \Omega_i^k} \cup_{\lambda_i \in \Lambda_i} \cup_{p_{\theta_i} \in P_i(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i)\} \\ &= \Pi_i^k(\theta'_i, \theta_i), \end{aligned}$$

as required. □

To shed further light on the role of ambiguous communication devices, we now consider the starkest possible case, whereby there is no ex-ante ambiguity. In those environments, all ambiguity is necessarily engineered by the mechanism designer.

### 4.3 No ex-ante ambiguity

Note that in environments with no ex-ante ambiguity (but with ambiguity-sensitive players), a social choice function is implementable by an unambiguous mechanism if and only if it is incentive compatible (with respect to the prior beliefs). Theorem 1 states, however, that when the designer can engineer some ambiguity in its mechanism, a larger set of social choice functions may be implemented. For instance, in the introductory example, the use of an ambiguous mechanism made it possible to implement an ex-post efficient allocation, while no unambiguous mechanism implements it. The following proposition gives a sharper characterization of condition (B) in Theorem 1 and, thus, to the scope of “belief engineering.”

**Proposition 1** *Assume that for each player  $i \in N$ , for each type  $\theta_i \in \Theta_i$ ,  $P_i(\theta_i)$  is the singleton  $\{p_{\theta_i}\}$ . Let  $\Pi_i^k$  be any non-empty finite-valued belief correspondence and denote  $\Pi_i^k(\theta'_i, \theta_i) = \{\pi_{\theta'_i, \theta_i}^{k, 1_k}, \dots, \pi_{\theta'_i, \theta_i}^{k, l_k}, \dots, \pi_{\theta'_i, \theta_i}^{k, L_k}\}$  the image of  $\Pi_i^k$  at  $(\theta'_i, \theta_i)$ . The following statements are equivalent:*

1. Condition (B) holds.

2. For each  $k \in K_i$ , for each  $\theta_i \in \Theta_i$  and  $\theta'_i \in \Theta_i$ ,  $|\Pi_i^k(\theta'_i, \theta_i)| = |\Pi_i^k(\theta_i, \theta'_i)|$ . For each  $\theta_i \in \Theta_i$ , for each  $\theta'_i \in \Theta_i$ , there exist positive scalars  $(\mu_{\theta'_i, \theta_i}^{k, l_k})$  such that  $\sum_k \sum_{l_k} \mu_{\theta'_i, \theta_i}^{k, l_k} = 1$ ,

$$p_{\theta_i} = \sum_{k=1}^{|K_i|} \sum_{l_k=1}^{L_k} \mu_{\theta'_i, \theta_i}^{k, l_k} \pi_{\theta'_i, \theta_i}^{k, l_k}, \quad (\text{B1})$$

and, for all  $\theta_{-i}$ ,

$$\mu_{\theta'_i, \theta_i}^{k, l_k} \pi_{\theta'_i, \theta_i}^{k, l_k}[\theta_{-i}] = \mu_{\theta'_i, \theta_i}^{k, l_k} \pi_{\theta'_i, \theta_i}^{k, l_k}[\theta_{-i}] \frac{p_{\theta_i}[\theta_{-i}]}{p_{\theta'_i}[\theta_{-i}]}. \quad (\text{B2})$$

To understand Proposition 1, suppose that the social choice function  $f$  is  $\Pi_i$ -incentive compatible. Condition (B1) states that the convex hull of the “targeted” beliefs  $\Pi_i(\theta_i, \theta_i)$  must include the prior beliefs  $p_{\theta_i}$  in its interior, for each  $\theta_i$ . This guarantees that one can engineer an ambiguity device such that the set of posterior beliefs of player  $i$  of type  $\theta_i$  is precisely  $\Pi_i(\theta_i, \theta_i)$ , regardless of the message received, provided that players truthfully report their types to the ambiguity device. This is a familiar result in the literature on repeated game with incomplete information (e.g., Aumann, Maschler and Stearns, 1995), known as the “splitting lemma.” It is worth stressing that condition (B1) does not require the prior to be a convex combination of each posterior in  $\Pi_i^k(\theta_i, \theta_i)$  for all  $k$ , but only to be a convex of the set of all posteriors (i.e.,  $\cup_k \Pi_i^k(\theta_i, \theta_i)$ ). Condition (B2) guarantees furthermore that the constructed ambiguity device generates the targeted beliefs  $\Pi_i(\theta'_i, \theta_i)$ , when player  $i$  of type  $\theta_i$  behave as type  $\theta'_i$  at the communication stage. Theorem 1 together with Proposition 1 are central results: they make it possible to “quantify” the role of endogenously engineered ambiguity in environments with no ex-ante ambiguity. For instance, Section 5 revisits the classical monopolistic screening problem and shows that full surplus extraction is possible, even when there is no ex-ante ambiguity.

**Proof** (1)  $\Rightarrow$  (2). Assume that condition (B) holds, so that there exist  $\Omega_i := \cup_{k \in K_i} \Omega_i^k$  and  $\lambda_i : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$  such that

$$\cup_{\omega_i \in \Omega_i^k} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i)\} = \Pi_i^k(\theta'_i, \theta_i),$$

for all  $\theta'_i \in \Theta_i$ , for all  $\theta_i \in \Theta_i$ , for all  $k \in K_i$ .

Clearly, we have that  $|\Pi_i^k(\theta'_i, \theta_i)| = |\Pi_i^k(\theta'_i, \theta'_i)|$  for any  $\theta_i$  and  $\theta'_i$ . Moreover, to any  $\pi_{\theta'_i, \theta_i}^{k, l_k}$  corresponds a message  $\omega_i \in \Omega_i^k$ , call it  $\omega_i^{k, l_k}$ , such that  $\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i^{k, l_k}, \lambda_i) = \pi_{\theta'_i, \theta_i}^{k, l_k}$ . Letting  $\mu_{\theta'_i, \theta_i}^{k, l_k} = (1/|K_i|) \sum_{\theta_{-i}} \lambda_i(\theta'_i, \theta_{-i})[\omega_i^{k, l_k}] p_{\theta_i}(\theta_{-i})$ , conditions (B1) and (B2) directly follow from the definition of  $\zeta_{\theta_i}$ .

(2)  $\Rightarrow$  (1). Let  $\Omega_i^k = \{\omega_i^{k, 1_k}, \dots, \omega_i^{k, l_k}, \dots, \omega_i^{k, L_k}\}$ . Construct  $\lambda_i : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\cup_{k \in K_i} \Omega_i^k)$  as follows: For each  $\theta_i \in \Theta_i$ , for each  $\theta_{-i} \in \Theta_{-i}$ ,

$$\lambda_i(\theta_i, \theta_{-i})[\omega_i^{k, l_k}] = \frac{1}{|K_i|} \frac{\mu_{\theta_i, \theta_i}^{k, l_k} \pi_{\theta_i, \theta_i}^{k, l_k}[\theta_{-i}]}{p_{\theta_i}[\theta_{-i}]}.$$

The rest directly follows from conditions (B1) and (B2) and the definition of  $\zeta_{\theta_i}$ .  $\square$

To conclude this section, we now consider environments with type-independent beliefs, a common assumption in applications.

#### 4.4 Type-independent Beliefs

It is common in applications to assume that prior beliefs are stochastically independent. This section considers a weaker notion of independence, type independence, that is  $P_i(\theta_i) = P_i$  for each  $\theta_i \in \Theta_i$ . In environments with type-independent beliefs, we have the following proposition.

**Proposition 2** *Consider environments with type-independent beliefs, i.e., for each player  $i \in N$ , for each  $\theta_i \in \Theta_i$ ,  $P_i(\theta_i) = P_i$ . The following statements are equivalent:*

1. *The social choice function  $f$  is implementable by a finite ambiguous mechanism.*
2. *For each player  $i \in N$ , there exists a finite family  $(\Pi_i^k)_{k \in K_i}$  of belief sets such that:*

(IC) *For each  $k \in K_i$ , the social choice function  $f$  is  $\Pi_i^k$ -incentive compatible.*

(B) *There exist  $\Omega_i := \cup_{k \in K_i} \Omega_i^k$  and  $\lambda_i : \Theta_i \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$  such that*

$$\cup_{p_i \in P_i} \cup_{\omega_i \in \Omega_i^k} \{\zeta_{\theta_i}(\theta'_i, p_i, \omega_i, \lambda_i)\} = \Pi_i^k, \quad (\text{B})$$

*for all  $\theta'_i \in \Theta_i$ , for all  $\theta_i \in \Theta_i$ , for all  $k \in K_i$ .*

According to Proposition 2, in environments with type-independent beliefs, we only need check the incentive compatibility of a social choice function with respect to sets of beliefs that are independent of the types of the players. The intuition is simple. Recall that  $\Pi_i(\theta'_i, \theta_i)$  is the set of beliefs of player  $i$  of type  $\theta_i$  at the allocation stage, when he communicates as type  $\theta'_i$  at the communication stage.<sup>15</sup> If beliefs are type-independent, we must have  $\Pi_i(\theta'_i, \theta_i) = \Pi_i(\theta'_i, \hat{\theta}_i)$  for any  $(\theta_i, \hat{\theta}_i)$ . Denote  $\Pi_i^*(\theta'_i)$  the set of beliefs when player  $i$  communicates as type  $\theta'_i$ . Therefore, the social choice function to be implemented must be incentive compatible with respect to  $\Pi_i^*(\theta'_i)$  for any  $\theta'_i$  and, thus, is incentive compatible with respect to  $\Pi_i^* := \Pi_i^*(\theta_i^*)$  for an arbitrary  $\theta_i^*$ .

Although substantially simpler than the characterization for general environments (Theorem 1), the above characterization is still not as simple and transparent as the characterization we have in environments with no ex-ante ambiguity. Clearly, a necessary condition for condition (B) to hold is that the set  $P_i$  of priors is a subset of the interior of the convex hull of  $\cup_k \Pi_i^k$  for each player  $i \in N$  (unless,  $\Pi_i^k = P_i$  for some  $k$ ). Conversely, if there are two players, two types per player and that each  $\Pi_i^k$  is a convex set, it is also a sufficient condition. Thus, in those simple environments, we have a relatively simple characterization. Proposition 3 summarizes this observation (the proof is in Appendix.)

**Proposition 3** *Consider environments with type-independent beliefs. Suppose that for each player  $i \in N$ , there exists a collection  $(\Pi_i^k)_k$  of beliefs' sets such the social choice function  $f$  is  $\Pi_i^k$ -incentive compatible for each  $k$ . Assume that there are two players, two types per player, and that each  $\Pi_i^k$  is a convex set, different from  $P_i$ .*

*A necessary and sufficient condition for condition (B) to hold is that the set  $P_i$  of priors is a subset of the interior of the convex hull of  $\cup_k \Pi_i^k$ , for each player  $i \in N$ .*

---

<sup>15</sup>I.e., we implicitly consider the canonical mechanism constructed in the proof of Theorem 1.

## 5 An Economic Application

This section revisits the classical problem of monopolistic screening, but with ambiguity-sensitive buyers. There are two potential buyers, labeled 1 and 2, and one monopolist seller. Buyer  $i$ 's utility from purchasing a quantity  $x_i$  of the good at the total price  $t_i$  is  $\theta_i x_i - t_i$ ;  $\theta_i$  can take two possible values  $\underline{\theta}$  and  $\bar{\theta}$  with  $\bar{\theta} > \underline{\theta} \geq 0$ . The reservation utility of a buyer is normalized to zero, regardless of his type. The seller's cost of producing the quantities  $(x_1, x_2)$  is  $c(x_1 + x_2)$  with  $c$  a strictly increasing and strictly convex function;  $c(0) = 0$ . For simplicity, we assume that  $c$  is differentiable everywhere and denote  $c'(x)$  the derivative of  $c$  at  $x$ . To rule out trivial corner solutions, let  $c'(0) = 0$ . Finally, let  $p \in (0, 1)$  be the (common) prior belief that  $\theta_i = \bar{\theta}$ .

So far, the problem is the classical textbook monopolistic screening problem. In particular, if the buyers are subjective expected maximizers, there is no mechanism, satisfying incentive compatibility and individual rationality, that allows the seller to extract full surplus from the buyers. Unlike the classical model, however, we assume that buyers have multiple-prior preferences.<sup>16</sup> Moreover, we allow the seller to design ambiguous (selling) mechanisms.

We first consider the first-best solutions. A first-best solution  $(x_1^*, x_2^*)$  is a solution to the following maximization program:

$$\max_{x_1, x_2} \theta_1 x_1 + \theta_2 x_2 - c(x_1 + x_2).$$

Standard arguments together with the strict convexity of  $c$  gives the following solution:

$$(x_1^*, x_2^*) = \begin{cases} ((1/2)(c')^{-1}(\underline{\theta}), (1/2)(c')^{-1}(\underline{\theta})) & \text{if } (\theta_1, \theta_2) = (\underline{\theta}, \underline{\theta}), \\ ((c')^{-1}(\bar{\theta}), 0) & \text{if } (\theta_1, \theta_2) = (\bar{\theta}, \underline{\theta}), \\ (0, (c')^{-1}(\bar{\theta})) & \text{if } (\theta_1, \theta_2) = (\underline{\theta}, \bar{\theta}), \\ ((1/2)(c')^{-1}(\bar{\theta}), (1/2)(c')^{-1}(\bar{\theta})) & \text{if } (\theta_1, \theta_2) = (\bar{\theta}, \bar{\theta}). \end{cases}$$

---

<sup>16</sup>We maintain the assumption of full Bayesian updating.

We now turn our attention to the optimal design problem between the seller and the buyers, when the seller can offer ambiguous “contracts.” From our revelation principle, we can restrict our attention to direct contracts  $(\theta_i, \theta_{-i}) \mapsto (x_i(\theta_i, \theta_{-i}), t_i(\theta_i, \theta_{-i}))$  and to finite families  $\Pi_i := \{\Pi_i^1, \dots, \Pi_i^k, \dots, \Pi_i^{K_i}\}$  of beliefs’ sets. (See Propositions 1 and 2.) Throughout, we write  $\Pi_i^k$  as  $\{\pi_i^{1,k}, \dots, \pi_i^{l,k}, \dots, \pi_i^{L_k,k}\}$ . The optimization program is thus:

$$\sup_{(x_1, x_2, t_1, t_2, \Pi_1, \Pi_2)} \sum_{(\theta_1, \theta_2)} (t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) - c(x_1(\theta_1, \theta_2) + x_2(\theta_1, \theta_2)))p[\theta_1]p[\theta_2]$$

subject to the following constraints:

**Incentive compatibility.** For each  $i \in \{1, 2\}$ , for each  $k \in \{1, \dots, K_i\}$ , for each  $\theta_i \in \{\underline{\theta}, \bar{\theta}\}$ ,

$$\begin{aligned} & \min_{\pi_i \in \Pi_i^k} \sum_{\theta_{-i}} (\theta_i x_i(\theta_i, \theta_{-i}) - t_i(\theta_i, \theta_{-i})) \pi_i[\theta_{-i}] \\ & \geq \min_{\pi_i \in \Pi_i^k} \sum_{\theta_{-i}} \sum_{\theta'_i} \sigma_i(\theta_i) [\theta'_i] (\theta_i x_i(\theta'_i, \theta_{-i}) - t_i(\theta'_i, \theta_{-i})) \pi_i[\theta_{-i}], \end{aligned} \tag{IC_i}$$

for all  $\sigma_i(\theta_i) \in \Delta(\{\underline{\theta}, \bar{\theta}\})$ ,

**Individual rationality.** For each  $i \in \{1, 2\}$ , for each  $\theta_i \in \{\underline{\theta}, \bar{\theta}\}$ ,

$$\sum_{\theta_{-i}} (\theta_i x_i(\theta_i, \theta_{-i}) - t_i(\theta_i, \theta_{-i})) p[\theta_{-i}] \geq 0, \tag{IR_i}$$

and

**Beliefs.** For each  $i \in \{1, 2\}$ , there exist  $(\mu_i^{l,k})_{l,k}$  with  $\mu_i^{l,k} > 0$  for all  $l \in \{1, \dots, L_k\}$ , for all  $k \in \{1, \dots, K_i\}$ , such that

$$p = \sum_{l,k} \mu_i^{l,k} \pi_i^{l,k}, \tag{B_i}$$



and

$$\sum_{l,k} \mu_i^{l,k} = 1.$$

Two remarks are worth making. Firstly, the individual rationality constraints ( $IR_i$ ) are defined ex-ante, since a buyer decides to participate (or not) in the mechanism before receiving any messages. Secondly, if we constraint  $\Pi_i$  to be  $\{\{p\}\}$  for each buyer  $i \in \{1, 2\}$  (equivalently, if the seller cannot offer ambiguous contracts), the problem boils down to the classical problem. Thus, if a solution exists, the seller is guaranteed to extract more surplus from the buyers than in the classical problem. In fact, we now show that the seller can extract *all* the surplus.

For any allocation  $x_1(\theta_1, \theta_2)$  and  $x_2(\theta_1, \theta_2)$ , ex-post full rent extraction obtains when no (informational) rents are left to the buyers, i.e.,  $t_1(\theta_1, \theta_2) = \theta_1 x_1(\theta_1, \theta_2)$  and  $t_2(\theta_1, \theta_2) = \theta_2 x_2(\theta_1, \theta_2)$ . If, in addition, the allocations are efficient, i.e.,  $x_1(\theta_1, \theta_2) = x_1^*(\theta_1, \theta_2)$  and  $x_2(\theta_1, \theta_2) = x_2^*(\theta_1, \theta_2)$ , ex-post full *surplus* extraction obtains. Let  $t_i^*$  be the corresponding payment scheme.

Clearly, if we can find a finite belief set  $\Pi_i^*$  that satisfies condition ( $B_i$ ) such that  $(x_i^*, t_i^*)$  is incentive compatible with respect to  $\Pi_i^*$ , then  $(x_i^*, t_i^*, \Pi_i^*)_{i \in \{1, 2\}}$  is a solution to the above problem.

So, consider the allocation corresponding to ex-post full surplus extraction, i.e.,  $(x_i^*, t_i^*)_{i \in \{1, 2\}}$ . The payoff to buyer  $i$  of type  $\bar{\theta}$  (resp.,  $\underline{\theta}$ ) is depicted in the left panel (resp., right panel) of the table below. For instance, if buyer  $i$  of type  $\bar{\theta}$  reports  $\underline{\theta}$  and buyer  $j$  also reports  $\underline{\theta}$ , then buyer  $i$ 's payoff is  $(1/2)(\bar{\theta} - \underline{\theta})(c')^{-1}(\underline{\theta})$ .

	$\bar{\theta}$	$\underline{\theta}$		$\bar{\theta}$	$\underline{\theta}$
$\bar{\theta}$	0	0	$\bar{\theta}$	$(\underline{\theta} - \bar{\theta}) \frac{(c')^{-1}(\bar{\theta})}{2}$	$(\underline{\theta} - \bar{\theta})(c')^{-1}(\bar{\theta})$
$\underline{\theta}$	0	$(\bar{\theta} - \underline{\theta}) \frac{(c')^{-1}(\underline{\theta})}{2}$	$\underline{\theta}$	0	0

Notice that truth-telling gives a payoff of zero to a buyer, regardless of his type and the report of the other buyer. Moreover, truth-telling is strictly dominant for a buyer

of type  $\underline{\theta}$ . Lastly, if  $\Pi_i^* = \{\{0, 1\}\}$ , then buyer  $i$  of type  $\bar{\theta}$  has an incentive to truthfully reveal his type when he expects the other buyer to do so. Since  $p \in (0, 1)$ , condition  $(B_i)$  is satisfied. Consequently,  $(x_i^*, t_i^*, \{\{0, 1\}\})_{i \in \{1, 2\}}$  is a solution to the seller's program. The solution is obviously not unique. However, it is easy to see that all solutions correspond to ex-post full surplus extraction.

The result of ex-post full surplus extraction generalizes to environments with more than two buyers and two types per buyer. To see this, suppose that there are  $n$  buyers and for each buyer  $i$ ,  $\Theta_i := \{\theta^1, \dots, \theta^l, \dots, \theta^L\}$ . Without loss of generality, assume that  $0 < \theta^1 < \dots < \theta^l \dots < \theta^L$ . We continue to assume type-independent and unambiguous prior beliefs, with  $p_i$  the prior beliefs of player  $i$  and  $p_i[\theta_{-i}] > 0$  for all  $\theta_{-i} \in \Theta_{-i}$ .

Let  $\theta_{-i}^{\max} := \max(\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$ . For ease of exposition, let the cost function be  $(1/2)(\sum_{i=1}^n x_i)^2$ . As before, the first-best allocation  $x^*$  is given by:  $x_i^*(\theta_i, \theta_{-i}) = \theta_i / |\{j : \theta_j \geq \theta_{-i}^{\max}\}|$  if  $\theta_i \geq \theta_{-i}^{\max}$  and  $x_i^*(\theta_i, \theta_{-i}) = 0$ , otherwise. Let  $t_i^*$  be the payment scheme such that  $t_i^*(\theta_i, \theta_{-i}) := \theta_i x_i^*(\theta_i, \theta_{-i})$ . Thus, the pair  $(x^*, t^*)$  corresponds to ex-post full surplus extraction.

The payoff to buyer  $i$  of type  $\theta^l$  as function of his report  $\theta^m$  and the highest report  $\theta_{-i}^{\max}$  of all other buyers is depicted in the table below ( $|\theta_{-i}^{\max} = \theta^m|$  denotes the cardinality of the set  $\{j \neq i : \theta_j = \theta_{-i}^{\max} = \theta^m\} \cup \{i\}$ ), i.e., the set of buyers reporting the highest valuations, conditional on buyer  $i$  reporting the highest valuation):

	$\theta^L$	...	$\theta^{l+1}$	$\theta^l$	$\theta^{l-1}$	...
$\theta^L$	$\frac{(\theta^l - \theta^L)\theta^L}{ \theta_{-i}^{\max} = \theta^L }$	$(\theta^l - \theta^L)\theta^L$	...			
$\vdots$						
$\theta^{l+1}$	0	0	$\frac{(\theta^l - \theta^{l+1})\theta^{l+1}}{ \theta_{-i}^{\max} = \theta^{l+1} }$	$(\theta^l - \theta^{l+1})\theta^{l+1}$	...	
$\theta^l$	0	0	0	0	0	0
$\theta^{l-1}$	0	0	0	0	$\frac{(\theta^l - \theta^{l-1})\theta^{l-1}}{ \theta_{-i}^{\max} = \theta^{l-1} }$	$(\theta^l - \theta^{l-1})\theta^{l-1}$
$\vdots$						

As above, it is immediate to verify that truth-telling is incentive compatible whenever

the set of beliefs of buyer  $i$  includes  $\pi_i^*$  with  $\pi_i^*[\theta_{-i}^{\max} = \theta^L] = 1$ . This is because the minimum expected payoff from truth-telling is zero (irrespective of beliefs about the other buyers), while the minimum expected payoff from any other strategy, mixed or pure, cannot be greater than the value of the expected payoff calculated using the belief  $\pi_i^*$ , which is at most zero. Lastly, since  $p_i$  is in the interior of the simplex  $\Delta(\Theta_{-i})$  and  $\pi_i^*$  is on the boundary of  $\Delta(\Theta_{-i})$ , we can find a set  $\Pi_i^*$  that includes  $\pi_i^*$  such that condition  $(B_i)$  holds.

With minor modifications, the same arguments apply to an optimal auction design problem (Myerson, 1981), so that ex-post full surplus extraction is possible.

As a final remark, consider the two-type model again and note that if buyer  $i$  has multiple priors  $[\underline{p}_i, \bar{p}_i]$  with  $0 < \underline{p}_i \leq \bar{p}_i < 1$ , then ex-post full surplus extraction remains the optimal solution. This follows from Proposition 3 and the observation that  $(x_i^*, t_i^*)_{i \in \{1,2\}}$  is incentive compatible with respect to  $[0, 1]$ .

## 6 Concluding Discussion

This section offers a critical discussion of some of the salient features of our analysis. Firstly, we argue that our results generalize without difficulties to a larger class of ambiguity-sensitive preferences.

### 6.1 Richer preferences

Most of our results generalize to a number of preferences having a “multiple-prior” representation. As a first example, consider the  $\alpha$ -maxmin criterion. It is clear that if we replace the incentive compatibility condition (IC) with the appropriate incentive condition, all our results go through. With  $\alpha$ -maxmin preferences, the appropriate incentive condition is:

$$\begin{aligned}
& \alpha \max_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) \\
& + (1 - \alpha) \min_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) \geq \\
& \alpha \max_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} \sum_{\theta'_i} \sigma_i(\theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) \\
& + (1 - \alpha) \min_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} \sum_{\theta'_i} \sigma_i(\theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}).
\end{aligned}$$

Variational preferences (Marinacci, Maccheroni and Rustichini, 2006) constitute another example. With the incentive condition

$$\begin{aligned}
& \min_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) + c_i(\pi_i, \theta_i) \geq \\
& \min_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} \sum_{\theta'_i} \sigma_i(\theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) + c_i(\pi_i, \theta_i),
\end{aligned}$$

most of our results go through.<sup>17</sup> Note, however, that Proposition 3 fails since we lose the linearity in beliefs.

Yet, another example is the criterion of minimax regret. With the incentive compatibility condition:

$$\begin{aligned}
& \max_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} [\max_{\tilde{\theta}_i} u_i(f(\tilde{\theta}_i, \theta_{-i}), \theta_i, \theta_{-i}) - u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i})] \pi_i(\theta_{-i}) \leq \\
& \max_{\pi_i \in \Pi_i(\theta_i'', \theta_i)} \sum_{\theta_{-i}} \sum_{\theta'_i} \sigma_i(\theta_i) [\theta'_i] [\max_{\tilde{\theta}_i} u_i(f(\tilde{\theta}_i, \theta_{-i}), \theta_i, \theta_{-i}) - u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i})] \pi_i(\theta_{-i}),
\end{aligned}$$

we have that any incentive compatible social choice function that satisfies condition (B) is implementable by an ambiguous mechanism (the second part of Theorem 1). However, the converse is not true. Simply, the criterion of minimax regret is not menu independent and, thus, the implementation of a social choice function by a given mechanism is not

---

<sup>17</sup>The function  $c_i(\cdot, \theta_i)$  is the ambiguity index of player  $i$  of type  $\theta_i$ .

equivalent to the implementation by the direct mechanism. See Saran (2011) for more on this issue.

As a final example, we consider the Bewley preferences (Bewley, 2002). Since those preferences are incomplete, there are two possible definitions of incentive compatibility (see Lopomo, Rigotti and Shannon (2010)). The first definition states that truth-telling dominates prior-by-prior any other report, whenever a player expects his opponents to tell the truth, that is,

$$\sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) \geq \sum_{\theta_{-i} \in \Theta_{-i}} u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}),$$

for all  $\theta'_i \in \Theta_i$ , for all  $\pi_i \in \Pi_i(\theta''_i, \theta_i)$ , for all  $\theta''_i \in \Theta_i$ .

The second definition simply states that truth-telling is not dominated and, thus, is implied by the first definition; that is, for all  $\theta_i \in \Theta_i$ , for all  $\theta''_i \in \Theta_i$ , no behavioral strategy  $\sigma_i(\theta''_i, \theta_i) \in \Delta(\Theta_i)$  satisfies

$$\sum_{\theta_{-i}} \sum_{\theta'_i} \sigma_i(\theta''_i, \theta_i) [\theta'_i] u_i(f(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}) > \sum_{\theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \pi_i(\theta_{-i}),$$

for all  $\pi_i \in \Pi_i(\theta''_i, \theta_i)$ . Nonetheless, regardless of the definition adopted, all our results go through. In particular, with the first definition, we have that a social choice function is implementable by an ambiguous mechanism if and only if it is implementable by a classic mechanism (i.e., unambiguous and static). In other words, a social choice function is implementable if and only if it is incentive compatible with respect to the prior beliefs.

Another important aspect of our analysis is the updating rule. We have assumed prior-by-prior updating (full Bayesian updating), a popular assumption in applications. However, other updating rules exist and, indeed, have axiomatic foundations. Following Gilboa and Schmeidler (1993), we can assume the maximum-likelihood updating rule. With that rule, a player updates the priors that maximize the likelihood of observing a particular event. We have the following:

**Proposition 4** *Assume the maximum-likelihood updating rule and no ex-ante uncertainty. Assume that, for each player  $i$ , there exists a finite-valued belief correspondence  $\Pi_i : \Theta_i \times \Theta_i \rightarrow \Delta(\Theta_{-i})$  such that the social choice function  $f$  is  $\Pi_i$ -incentive compatible with  $p_{\theta_i}$  the arithmetic mean of the elements of  $\Pi_i(\theta_i'', \theta_i)$  for all  $\theta_i''$ . Then, the social choice function  $f$  is implementable by an ambiguous mechanism.*

The intuition is simple. Consider the construction of the ambiguous communication device in Proposition 1. If priors are arithmetic means of the targeted posteriors, we have that each of the probability systems is equally likely to have generated a given message, so that maximum-likelihood updating coincides with prior-by-prior updating. Note, however, that the introductory example did not satisfy the above proposition.

Another updating rule we mention is Hanany-Klibanoff updating rules. Because these updating rules guarantee a weak form of dynamic consistency, we generally have the classical revelation principle (adapted to the preferences, of course). To be more precise, suppose that  $f$  is not incentive compatible, i.e., there exist player  $i$ , types  $\theta_i$  and  $\theta_i^*$  such that  $f_{\theta_i^*} \succ_i^{\theta_i} f_{\theta_i}$ , where  $f_{\theta_i}$  denotes the act that gives  $f(\theta_i, \theta_{-i})$  in state  $\theta_{-i}$  and  $\succ_i^{\theta_i}$  the strict preference of player  $i$  of type  $\theta_i$ . Consider the set of all such acts, i.e.,  $F := \cup_{\theta_i' \in \Theta_i} \{f_{\theta_i'}\}$ . Since the set of types is finite, the set  $F$  is finite and, consequently, we can assume that  $f_{\theta_i^*} \succ_i^{\theta_i} f_{\theta_i'}$  for all  $\theta_i'$ . So, consider the set of acts  $F$ , the ex-ante optimal act  $f_{\theta_i^*}$ , the ex-ante preferences  $\succ_i^{\theta_i}$  and the event  $\{\omega_i\}$ , corresponding to a profile of messages received and sent by player  $i$  before the allocation stage. Moreover, since the social choice function does not depend on cheap talk messages, we have that two acts in  $F$  agree outside  $\{\omega_i\}$ . Consequently, dynamic consistency as defined in Hanany and Klibanoff requires  $f_{\theta_i^*} \succ_{\{\omega_i\}, f_{\theta_i^*}, F}^{\theta_i} f_{\theta_i}$ , while implementation requires  $f_{\theta_i} \succ_{\{\omega_i\}, f_{\theta_i^*}, F}^{\theta_i} f_{\theta_i^*}$  ( $\succ_{\{\omega_i\}, f_{\theta_i^*}, F}^{\theta_i}$  is player  $i$ 's conditional preference). This violates axiom **DC3** (p. 273) in Hanany and Klibanoff, but not their main axiom **DC** (p. 270) of dynamic consistency. As already explained, the introductory example can be modified so that their main axiom is satisfied. Consequently, a violation of their axiom **DC** is not a necessary condition for our results to hold.

Of course, this discussion does not exhaust all possible updating rules and ambiguity-sensitive preferences. It simply suggests that our results extend to a larger class of ambiguity-sensitive preferences and updating rules. Further research is certainly needed.

## 6.2 Ambiguity In The Allocation Mechanism

A noteworthy feature of ambiguous mechanisms is that no ambiguity is introduced at the allocation stage; ambiguity is only introduced at the communication stage. This is, however, without loss of generality. To see this, suppose that the allocation mechanism is  $\langle (M_i)_{i \in N}, G \rangle$  with  $G$  a set of allocation rules  $g : M \rightarrow X$ , so that players are ambiguous about the exact allocation rule used for implementation. For instance, the alternative implemented might depend on the messages sent by the players as well as a draw from an urn, which contains either blue balls or red balls, but with unknown proportions (an “Ellsberg” urn.)

For the social choice function  $f$  to be implementable, however, there must exist a consistent planning equilibrium  $s^*$  such that  $g(m_i, m_{-i}) = f(\theta_i, \theta_{-i})$  for all  $(m_i, m_{-i})$  with  $\times_{i \in N} s_i^*(\theta_i, h_i^{T+1})[m_i] > 0$ , for all  $(h_i^{T+1}, h_{-i}^{T+1})$  with positive probability under  $s^*$ , and for all  $g \in G$ .<sup>18</sup> Consequently, introducing ambiguity in the allocation mechanism does not help in the partial implementation of social choice functions. This observation may explain why most real-world mechanisms, e.g., voting systems, schooling allocation mechanisms, auctions, have unambiguous allocation rules. Yet, as the work of Di Tillio, Kos and Messmer suggests, ambiguity at the allocation stage might enlarge the set of implementable social choice sets.

---

<sup>18</sup>To see this, note that if there exist  $g \in G$ ,  $(h_i^{T+1}, h_{-i}^{T+1})$  with positive probability under  $s^*$ ,  $(m_i, m_{-i})$  with  $\times_{i \in N} s_i^*(\theta_i, h_i^{T+1})[m_i] > 0$  such that  $g(m_i, m_{-i}) \neq f(\theta_i, \theta_{-i})$ , then we have a violation of the definition of implementation.

### 6.3 Beyond Ambiguous Mechanisms

Ambiguous mechanisms are relatively simple. They consist of a finite number of stages of mediated communication followed by an allocation stage. Clearly, there are more general mechanisms. For instance, the communication device at stage  $t$  might depend on the entire histories of messages sent and received up to period  $t$ . Similarly, for the allocation. The following example illustrates some of the challenges in obtaining a revelation principle for general ambiguous mechanisms.

There are two players, labeled 1 and 2, two types  $\theta$  and  $\theta'$  for each player, and two alternatives  $x$  and  $y$ . Types are private information. We assume that players have multiple-prior preferences (Gilboa and Schmeidler, 1989) with  $P_i$  the set of priors of player  $i \in \{1, 2\}$  and  $u_i$  his utility function. Suppose that  $u_1(x, \theta) = 1$ ,  $u_1(y, \theta) = 0$ , that player 1 of type  $\theta'$  and player 2 of both types are indifferent between all alternatives, and that  $P_i$  is independent of player  $i$ 's type. An element of  $P_i$  represents a prior belief of player  $i$  about the likelihood of player  $j$ 's type to be  $\theta$ . For concreteness, assume that  $\{p_i\} = \{1/3\}$ .<sup>19</sup>

The designer aims at (partially) implementing the social choice function  $f$  defined by:  $f(\theta, \theta) = x$ ,  $f(\theta, \theta') = y$ ,  $f(\theta', \theta) = y$  and  $f(\theta', \theta') = x$ .

The social choice function  $f$  is not implementable by an ambiguous mechanism. Regardless of his beliefs, player 1 can guarantee a payoff of  $1/2$  by mixing uniformly between  $\theta$  and  $\theta'$ . So, to satisfy the incentive compatibility constraints (IC), we need to find a finite collection of finite beliefs' sets  $(\Pi_i^k)_k$  such that  $\min \Pi_i^k \geq 1/2$  for each  $k$ . However, to satisfy condition (B), we also need that the prior belief  $1/3$  belongs to the convex hull of  $\cup_k \Pi_i^k$ , which is impossible.

Yet, we claim that the social choice function is implementable by a more general ambiguous mechanism. The mechanism has three stages. In the first stage, player 2

---

<sup>19</sup>More generally, all our arguments remain valid if  $P_i = [\underline{p}_i, \bar{p}_i]$  with  $\underline{p}_i < 1 - \bar{p}_i$ . Alternatively, if  $\underline{p}_i \geq 1 - \bar{p}_i$ ,  $f$  is implementable with a classical direct mechanism, so that there is no need for non-trivial ambiguous mechanisms.



reports either  $\theta$  or  $\theta'$  to the designer. Following player 2's report, the designer sends either  $\omega$  or  $\omega'$  to player 1. There are two possible probability systems,  $\lambda$  and  $\lambda'$ . The first probability system  $\lambda$  is fully specified by  $\lambda(\omega|\theta) = 1$  and  $\lambda(\omega'|\theta') = 1$ , while  $\lambda'(\omega'|\theta) = 1$  and  $\lambda'(\omega|\theta') = 1$  fully specify the second probability system. Player 1 is not active in the first stage. In the second stage, player 1 reports either  $\theta$  or  $\theta'$  to the designer. If player 1 reports  $\theta$ , the designer implements  $f(\theta, \theta)$  (resp.,  $f(\theta, \theta')$ ) if player 2 reported  $\theta$  (resp.,  $\theta'$ ) in the first stage. Alternatively, if player 1 reports  $\theta'$ , the mechanism moves to the third and final stage. In the third stage, player 1 has again to report  $\theta$  or  $\theta'$ . If player 1 reports  $\theta'$ , the designer implements  $f(\theta', \theta)$  (resp.,  $f(\theta', \theta')$ ) if player 2 reported  $\theta$  (resp.,  $\theta'$ ) in the first stage. Alternatively, if player 1 reports  $\theta$ , the designer implements  $y$ , regardless of player 2's report. (Player 2 is not active at the second and third stage.) The distinctive feature of this mechanism is the multi-stage allocation mechanism. See Figure 3 for a graphical illustration.

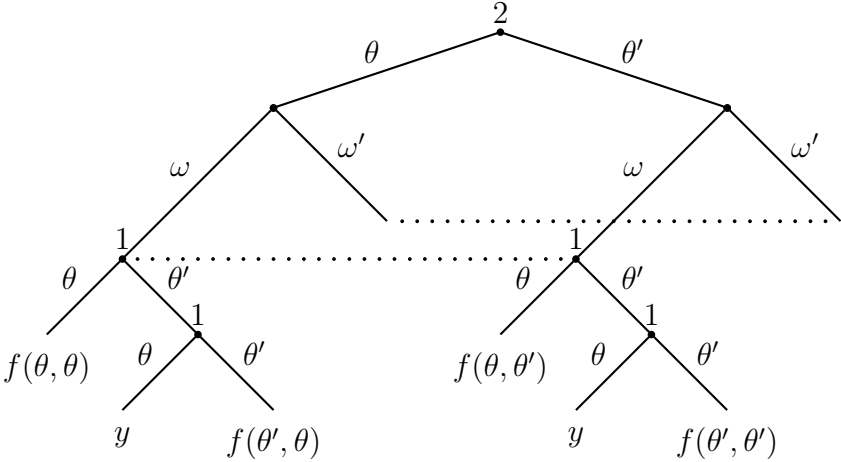


Figure 3: The mechanism

We now argue that both players have an incentive to truthfully reveal their types at all stages. Since player 1 of type  $\theta'$  and player 2 of either type are indifferent between all alternatives, they clearly have an incentive to truthfully reveal their types. So, let us focus on player 1 of type  $\theta$ . Consider the history  $(\omega, \theta')$ , i.e., player 1 has received the message  $\omega$  from the designer at the first stage and has reported  $\theta'$  at the second stage. By

construction of the ambiguous communication device, player 1's set of beliefs is  $\{0, 1\}$ , i.e., he believes that either player 2 is of type  $\theta$  with probability 1 or of type  $\theta'$  with probability 1. At  $(\omega, \theta')$ , player 1 is indifferent between reporting  $\theta$ , which guarantees a payoff of zero and reporting  $\theta'$ . Moreover, no mixture between  $\theta$  and  $\theta'$  is strictly preferred to reporting  $\theta$ .<sup>20</sup> Consequently, it is optimal for player 1 of type  $\theta$  to truthfully report  $\theta$  at the third stage following the history  $(\omega, \theta')$ . Let us move to the history  $(\omega)$ . At  $(\omega)$ , player 1's set of beliefs is  $\{0, 1\}$  and so he is indifferent between reporting  $\theta$  and any mixing between  $\theta$  and  $\theta'$  (conditional on reporting  $\theta$  at the third stage and, thus, obtaining  $y$  for sure). So, it is optimal for player 1 of type  $\theta$  to truthfully report  $\theta$  at  $\omega$ . A similar argument holds at  $\omega'$ , so that  $f$  is indeed implementable by the constructed mechanism.

## 6.4 Continuum

This section generalizes our main result to environments and mechanisms with a continuum of alternatives, types, and messages. In particular, even in environments with a finite number of alternatives and types, the designer may benefit from using a continuum of messages. It makes it possible to engineer larger sets of beliefs and, consequently, to relax the incentive-compatibility constraints.

Throughout, we endow any metrizable space  $Y$  with  $\mathcal{B}_Y$ , the  $\sigma$ -algebra of Borel sets on  $Y$ , to form the probability space  $(Y, \mathcal{B}_Y)$ . Denote  $\Delta(Y)$  the set of all probability measures on  $(Y, \mathcal{B}_Y)$ . We endow  $\Delta(Y)$  with the weak\* topology. Let  $Y$  and  $Y'$  be two metrizable spaces. A function  $\phi$  from  $Y$  to  $\Delta(Y')$  is measurable if  $\phi^{-1}(O_{Y'}) \in \mathcal{B}_Y$  for all  $O_{Y'} \in \mathcal{B}_{\Delta(Y')}$ , the Borel  $\sigma$ -algebra on  $\Delta(Y')$  endowed with the weak\* topology. A probability measure  $\mu$  on  $(Y, \mathcal{B}_Y)$  admits the probability density  $\zeta$  if  $\mu$  is absolutely continuous with respect to the Lebesgue measure on  $Y$  and has Radon-Nikodym derivative  $\zeta$  (with respect to the Lebesgue measure). We denote  $\hat{\Delta}(Y) \subseteq \Delta(Y)$  the set of such probability

---

<sup>20</sup>This follows from the  $c$ -independence of the multiple prior preferences (Gilboa and Schmeidler, 1989).

measures. With a slight abuse of notation, we write  $\hat{\Delta}(Y)$  for the set of probability densities corresponding to the set of measures  $\hat{\Delta}(Y)$ . Finally, we endow products of topological spaces with the product topology.

With these mathematical preliminaries done, assume that all sets introduced in Section 3 (i.e.,  $X$ ,  $(\Theta_i)_i$ ,  $(\Omega_{i,t}, \Omega_{i,t}^*)_{i,t}$ ,  $(M_i)_i$ , etc) are subsets of metrizable spaces; that functions (i.e., payoff functions, probability systems, strategies, etc) are measurable, and that each probability measure admits a strictly positive density. For each  $t$ ,  $\Lambda_t$  is a measurable subset of the set of all measurable probability systems  $\lambda_t : \Omega_t \rightarrow \Delta(\Omega_t^*)$  with densities. Assume that each  $\{\lambda_t\}$  is measurable and that there exists a density  $\nu$  such that  $\nu(\lambda_t) > 0$ .

With these technical assumptions, we have that

$$\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda)[\theta_{-i}] = \frac{\int_{\omega_{-i}} \hat{\lambda}((\theta'_i, \theta_{-i}))[(\omega_i, \omega_{-i})] \hat{p}_{\theta_i}(\theta_{-i}) d\omega_{-i}}{\int_{\omega_{-i}} \int_{\theta_{-i}} \hat{\lambda}((\theta'_i, \theta_{-i}))[(\omega_i, \omega_{-i})] \hat{p}_{\theta_i}(\theta_{-i}) d\omega_{-i} d\theta_{-i}},$$

with  $\hat{\lambda}(\theta)$  the probability density corresponding to the measure  $\lambda(\theta)$  and  $\hat{p}_{\theta_i}$  the probability density corresponding to the measure  $p_{\theta_i}$ .

**Theorem 2** *The following statements are equivalent:*

1. *The social choice function  $f$  is implementable by an ambiguous mechanism.*
2. *For each player  $i \in N$ , there exists a family of non-empty valued correspondences  $(\Pi_i^k : \Theta_i \times \Theta_i \rightarrow \hat{\Delta}(\Theta_{-i}))_{k \in K_i}$  such that:*

(IC) *For each  $k \in K_i$ , the social choice function  $f$  is  $\Pi_i^k$ -incentive compatible.*

(B) *There exist  $\Omega_i := \cup_{k \in K_i} \Omega_i^k$  and  $\lambda_i : \Theta_i \times \Theta_{-i} \rightarrow \hat{\Delta}(\Omega_i)$  such that*

$$\cup_{p_{\theta_i} \in P(\theta_i)} \cup_{\omega_i \in \Omega_i^k} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i)\} = \Pi_i^k(\theta'_i, \theta_i), \quad (\text{B3})$$

*for all  $\theta'_i \in \Theta_i$ , for all  $\theta_i \in \Theta_i$ , for all  $k \in K_i$ .*

**Proof** Only minor modifications to the proof of Theorem 1 are required. The only modification required to prove that (2)  $\Rightarrow$  (1) is to show that the subgroup of cyclic permutations from  $\Omega_i$  to  $\Omega_i$  has the cardinality of  $\Omega_i$ . Let  $\rho_{\alpha_0} : \Omega_i \rightarrow \Omega_i$  be the identity function and define inductively the bijection  $\rho_\alpha : \Omega_i \rightarrow \Omega_i$  such that  $\rho_\alpha(\omega_i) \neq \rho_{\alpha'}(\omega_i)$  for all  $\alpha' < \alpha$ , for all  $\omega_i \in \Omega$ . Let  $A$  be the set of all such  $\alpha$ . Clearly, the cardinality of  $A$  is weakly smaller than the cardinality of  $\Omega_i$ , so suppose that  $|A| < |\Omega_i|$ . It follows that  $|\cup_\alpha \{\rho_\alpha(\omega_i)\}| = |A| < |\Omega_i|$ . From the axiom of choice, it follows that for each  $\omega_i \in \Omega_i$ , we can choose  $\rho(\omega_i) \in \Omega_i \setminus (\cup_\alpha \{\rho_\alpha(\omega_i)\})$  such that  $\rho$  is a bijection, a contradiction.

The modification required to prove that (1)  $\Rightarrow$  (2) is in the definition of  $\lambda_i^*$ : we define  $\lambda_i^*(\theta_i, \theta_{-i})[(\omega_i, \lambda_i)] := \lambda_i(\theta_i, \theta_{-i})[\omega_i]\nu(\lambda_i)$ , with  $\nu$  the density on  $\Lambda_i$ .  $\square$

## Appendix

**Proof of Theorem 1: the general case** We treat the case of a family of correspondences. The proof is almost identical to the proof in the main text. We only sketch the important differences.

Suppose that for each player  $i \in N$ , there exists a family of non-empty valued correspondences  $\Pi_i^k : \Theta_i \times \Theta_i \rightarrow \Delta(\Theta_{-i})$  such that conditions (IC) and (B) hold. Since condition (B) holds, there exists  $(\Omega_i^k)_{k \in K_i}$  and  $\lambda_i$  such that  $\cup_{p_{\theta_i} \in P(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i)\} \subseteq \Pi_i^k(\theta'_i, \theta_i)$  for all  $\omega_i \in \Omega_i^k$ .

For each  $k \in K_i$ , consider a permutation  $\rho^k : \Omega_i^k \rightarrow \Omega_i^k$  and the probability system  $\lambda_i^\rho : \Theta \times \Theta_{-i} \rightarrow \Delta(\Omega_i)$  defined by  $\lambda_i^\rho(\theta)[\omega_i] = \lambda_i(\theta)[\rho^k(\omega_i)]$  for all  $\omega_i \in \Omega_i^k$ , for all  $k \in K_i$ . Define  $\Lambda_i$  as in the proof of Theorem 1 (i.e., as  $\{\lambda^\rho : \rho^k \text{ is a cyclic permutation for each } k\}$ ), we obtain that  $\cup_{\lambda_i \in \Lambda_i} \cup_{p_{\theta_i} \in P(\theta_i)} \{\zeta_{\theta_i}(\theta'_i, p_{\theta_i}, \omega_i, \lambda_i)\} = \Pi_i^k(\theta'_i, \theta_i)$  for all  $\omega_i \in \Omega_i^k$ , for all  $k \in K_i$ . The rest of the proof is almost identical to the proof of Theorem 1 and left to the reader.  $\square$

**Lemma 1** *If the social choice function  $f$  is implementable by the ambiguous mechanism*

$$\langle\langle(\Omega_{i,t}^*, \Omega_{i,t})_{i \in N}, \Lambda_t\rangle_{t=1, \dots, T}, \langle\langle M_i\rangle_{i \in N}, g\rangle\rangle,$$

*then  $f$  is implementable by a two-stage ambiguous mechanism.*

**Proof of Lemma 1** Suppose that the social choice function  $f$  is implementable by the ambiguous mechanism

$$\langle\langle(\Omega_{i,t}^*, \Omega_{i,t})_{i \in N}, \Lambda_t\rangle_{t=1, \dots, T}, \langle\langle M_i\rangle_{i \in N}, g\rangle\rangle,$$

and let  $(s^*, \Pi^{H, \Theta})$  be the consistent planning equilibrium implementing  $f$ . Consider the two-stage mechanisms:

$$\langle\langle(\Theta_i, \times_t(\Omega_{i,t}^* \times \Omega_{i,t}))_{i \in N}, \Lambda\rangle, \langle\langle M_i\rangle_{i \in N}, g\rangle\rangle,$$

with  $\Lambda$  constructed as follows. For any  $(\lambda_1, \dots, \lambda_T) \in \Lambda_1 \times \dots \times \Lambda_T$ , we associate the communication system  $\lambda$  defined by:

$$\begin{aligned} \lambda(\theta)[(\omega_t^*, \omega_t)_{t=1}^T] &:= s^*(\theta)[\omega_1^*] \lambda_1(\omega_1^*)[\omega_1] \times \dots \\ &\times \dots \times s^*(\theta, (\omega_t^*, \omega_t)_{t=1}^T)[\omega_\tau^*] \lambda_\tau(\omega_\tau^*)[\omega_\tau] \times \dots \\ &\times \dots \times s^*(\theta, (\omega_t^*, \omega_t)_{t=1}^{T-1})[\omega_T^*] \lambda_T(\omega_T^*)[\omega_T]. \end{aligned}$$

Let  $H^{**}$  the relevant histories. Note each element of  $H^{**}$  is an element of  $H$ . Consider the profile of strategies  $s^{**}$  with  $s_i^{**}(\theta_i) = \theta_i$  at the initial history and  $s_i^{**}(\theta_i, (\omega_{i,t}^*, \omega_{i,t})_{t=1}^T) = s_i^*(\theta_i, h_i^{T+1})$  with  $h_i^{T+1} = ((\omega_{i,t}^*, \omega_{i,t})_{t=1}^T)$ .

By construction, if the history  $h_i^{T+1} = ((\omega_{i,t}^*, \omega_{i,t})_{t=1}^T)$  has positive probability under  $s^*$ , then the profile of messages  $((\omega_{i,t}^*, \omega_{i,t})_{t=1}^T)$  has positive probability under  $s^{**}$ . Moreover, the set of beliefs of player  $i$  of type  $\theta_i$  conditional on the message  $((\omega_{i,t}^*, \omega_{i,t})_{t=1}^T)$  is exactly equal to  $\Pi_i^{H, \Theta}(\theta_i, h_i^{T+1})$ .

Since the consistent planning equilibrium  $(s^*, \Pi^{H, \Theta})$  implements  $f$ , it follows that the strategy profile  $(s^{**}, \Pi^{H^{**}, \Theta})$  is a consistent planning equilibrium of the game induced by the two-stage mechanism and, furthermore, implements  $f$ .  $\square$

**Proof of Proposition 2** (2)  $\Rightarrow$  (1). This follows from Theorem 1.

(1)  $\Rightarrow$  (2). The proof is almost identical to the proof of Theorem 1. We only sketch the main differences. Since beliefs are type-independent, we have that  $\Pi^\Theta(\theta_i, (\theta_i, \omega_i)) = \Pi_i^\Theta(\theta'_i, (\theta_i, \omega_i))$  for any  $\theta_i$  and  $\theta'_i$ , for any history  $(\theta_i, \omega_i)$ . Denote  $\Pi^\Theta((\theta_i, \omega_i))$  the beliefs at history  $(\theta_i, \omega_i)$ .

For each player  $i \in N$ , fix a type  $\theta_i^* \in \Theta_i$  and for each  $\lambda \in \Lambda$ , define  $\lambda_i^* : \Theta_{-i} \rightarrow \Delta(\Omega_i)$  with  $\lambda_i^*(\theta_{-i})[\omega_i] = \sum_{\omega_{-i}} \lambda(\theta_i^*, \theta_{-i})[(\omega_i, \omega_{-i})]$ . Denote  $\Lambda_i^*$  the set of such probability systems.

We claim the social choice function  $f$  is implementable by

$$\langle\langle (\Theta_i, \Omega_i, \Lambda_i^*)_{i \in N} \rangle, \langle (M_i)_{i \in N}, g \rangle\rangle.$$

Consider the strategy profile  $s^*$  for which players truthfully reports their types at the first stage and at the second stage, regardless of the history. It is a consistent planning equilibrium. To see this, notice that regardless of player  $i$ 's report at the first stage, the set of player  $i$ 's beliefs at the history  $(\theta_i, \omega_i)$  is  $\Pi_i^\Theta((\theta_i^*, \omega_i))$ . Moreover, since  $f$  is implementable by the mechanism

$$\langle\langle (\Theta_i, \Omega_i)_{i \in N}, \Lambda \rangle, \langle (M_i)_{i \in N}, g \rangle\rangle,$$

we have from Theorem 1 that  $f$  is incentive compatible with respect to  $\Pi_i^\Theta((\theta_i^*, \omega_i))$  for each player  $i \in N$ , for each  $\omega_i$ . Thus,  $s^*$  is indeed a consistent planning equilibrium. The rest of the proof then follows as in the proof of Theorem 1.  $\square$

**Proof of Proposition 3** Suppose that there are two players, labeled 1 and 2, and each player  $i \in \{1, 2\}$  has two types  $\theta$  and  $\theta'$ . Let  $P_i \subseteq \Delta(\{\theta, \theta'\})$  be the non-empty, convex, closed set of priors of player  $i$ . Suppose that  $f$  is incentive compatible with respect to the non-empty, convex and closed set of beliefs  $\Pi_i$  for each player  $i$ . Notice that since player  $i$ 's expected payoff is linear in beliefs,  $f$  is also incentive compatible with respect to  $\Pi_i^*$ , with  $\{\pi_i^1, \pi_i^2\} \subseteq \Pi_i^* \subseteq \Pi_i$  and  $\{\pi_i^1, \pi_i^2\}$  the set of extreme points of  $\Pi_i$ , for each player  $i$ . Let  $p_i$  and  $p'_i$  be the two extreme points of  $P_i$  with  $p_i(\theta) \leq p'_i(\theta)$ . Without loss of generality, assume that  $\pi_i^1(\theta) \leq \pi_i^2(\theta)$ .

Assume that  $p_i(\theta)$  and  $p'_i(\theta)$  are strictly in between  $\pi_i^1(\theta)$  and  $\pi_i^2(\theta)$ , so that  $P_i$  is in the interior of  $\Pi_i$ .

Fix  $q_i^2 \in \Delta(\{\theta, \theta'\})$  such that  $p_i = \mu_1 \pi_i^1 + \mu_2 q_i^2$  with  $\mu_1 > 0$ ,  $\mu_2 > 0$  and  $\mu_1 + \mu_2 = 1$ . Let  $\Omega = \{\omega_1, \omega_2\}$ . The idea is to construct a probability system  $\lambda_i : \{\theta, \theta'\} \rightarrow \Delta(\Omega)$  such that player  $i$ 's posterior is  $\pi_i^1$  (resp.,  $q_i^2$ ) conditional on  $\omega_1$  (resp.,  $\omega_2$ ), when he uses the prior  $p_i$ , and to “control”  $q_i^2$  so that player  $i$ 's posterior is  $\pi_i^2$  conditional on  $\omega_2$ , when he uses the prior  $p'_i$ . Define  $\lambda_i$  as follows:

$$\begin{aligned}\lambda_i(\theta)[\omega_1] &= \frac{\mu_1 \pi_i^1(\theta)}{p_i(\theta)}, \\ \lambda_i(\theta)[\omega_2] &= \frac{\mu_2 q_i^2(\theta)}{p_i(\theta)}, \\ \lambda_i(\theta')[\omega_1] &= \frac{\mu_1 \pi_i^1(\theta')}{p_i(\theta')}, \\ \lambda_i(\theta')[\omega_2] &= \frac{\mu_2 q_i^2(\theta')}{p_i(\theta')}.\end{aligned}$$

Denote  $\zeta_i(p_i, \omega, \lambda_i)$  the posterior belief of player  $i$  about the type of his opponent, conditional on the message  $\omega$ , when he uses the prior  $p_i$ . By construction, we have that  $\zeta_i(p_i, \omega_1, \lambda_i) = \pi_i^1$  and  $\zeta_i(p_i, \omega_2, \lambda_i) = q_i^2$ . We now derive  $\zeta_i(p'_i, \omega, \lambda_i)$  for any  $\omega \in \Omega$ . We have

$$\begin{aligned}\zeta_i(p'_i, \omega_1, \lambda_i)[\theta] &= \frac{\pi_i^1(\theta)(p'_i(\theta)/p_i(\theta))}{\pi_i^1(\theta)(p'_i(\theta)/p_i(\theta)) + \pi_i^1(\theta')(p'_i(\theta')/p_i(\theta'))} := q_i^1(\theta), \\ \zeta_i(p'_i, \omega_2, \lambda_i)[\theta] &= \frac{q_i^2(\theta)(p'_i(\theta)/p_i(\theta))}{q_i^2(\theta)(p'_i(\theta)/p_i(\theta)) + q_i^2(\theta')(p'_i(\theta')/p_i(\theta'))}, \\ \zeta_i(p'_i, \omega_1, \lambda_i)[\theta'] &= 1 - \zeta_i(p'_i, \omega_1, \lambda_i)[\theta], \\ \zeta_i(p'_i, \omega_2, \lambda_i)[\theta'] &= 1 - \zeta_i(p'_i, \omega_2, \lambda_i)[\theta].\end{aligned}$$

So, we want to find  $q_i^2$  such that  $\zeta_i(p'_i, \omega_2, \lambda_i) = \pi_i^2$ , while  $q_i^1$  must belong to  $\Pi_i$ .

Notice that the function  $\phi : [0, 1] \rightarrow \mathbb{R}_+$ , defined by

$$\phi(y) = \frac{y(p'_i(\theta)/p_i(\theta))}{y(p'_i(\theta)/p_i(\theta)) + (1-y)(p'_i(\theta')/p_i(\theta'))},$$

is continuous and increasing in  $y \in [0, 1]$ . Moreover,

$$\frac{y(p'_i(\theta)/p_i(\theta))}{y(p'_i(\theta)/p_i(\theta)) + (1-y)(p'_i(\theta')/p_i(\theta'))} = y \frac{1}{y + (1-y) \frac{p'_i(\theta') p_i(\theta)}{p_i(\theta') p'_i(\theta)}} \geq y,$$

since  $p_i(\theta) \leq p'_i(\theta)$ ,  $p_i(\theta') = 1 - p_i(\theta)$ , and  $p'_i(\theta') = 1 - p'_i(\theta)$ .

For a solution to exist, we need to satisfy the following inequalities:

$$\begin{aligned} \pi_i^1(\theta) &\leq q_i^1(\theta) = \phi(\pi_i^1(\theta)) \leq \pi_i^2(\theta), \\ \phi(\pi_i^1(\theta)) &\leq \pi_i^2(\theta), \\ q_i^2(\theta) &\geq p_i(\theta). \end{aligned}$$

The first inequality is required for  $q_i^2$  to be in the convex hull of  $\{\pi_i^1, \pi_i^2\}$ , the second inequality is required for the existence of a solution of  $f(q_i^2) = \pi_i^2$ , and the last inequality is required for the  $\mu$ 's to be well-defined.

Notice that first inequality is equivalent to the second inequality. Moreover, the last inequality is automatically satisfied if the second is. To see this, observe that the second inequality implies that there exists  $q_i^2$  such that  $\phi(q_i^2(\theta)) = \pi_i^2(\theta)$ ; this follows from the intermediate value theorem. By contradiction, assume that  $q_i^2(\theta) < p_i(\theta)$ . Since  $\phi$  is increasing, we have that  $\phi(q_i^2(\theta)) = \pi_i^2(\theta) \leq \phi(p_i(\theta)) = p'_i(\theta) < \pi_i^2(\theta)$ , a contradiction.

From the definition of  $\phi$ , it follows that a solution exists, whenever

$$\frac{\pi_i^1(\theta) \pi_i^2(\theta')}{\pi_i^1(\theta') \pi_i^2(\theta)} \leq \frac{(p'_i(\theta')/p_i(\theta'))}{(p'_i(\theta)/p_i(\theta))} = \frac{p'_i(\theta') p_i(\theta)}{p_i(\theta') p'_i(\theta)}.$$

The last inequality is satisfied whenever  $p_i(\theta)$  and  $p'_i(\theta)$  are in between  $\pi_i^1(\theta)$  and  $\pi_i^2(\theta)$ , as postulated. To summarize, we have that  $\cup_{\tilde{p}_i \in P_i} \cup_{\omega \in \Omega} \{\zeta_i(\tilde{p}_i, \omega, \lambda_i)\} = \text{co}\{\pi_i^1, q_i^2\} \cup \text{co}\{q_i^1, \pi_i^2\} := \Pi_i^* \subseteq \Pi_i$ , as required.  $\square$

## References

- [1] Al-Najjar, Nabil and Jonathan Lewis Weinstein. "The Ambiguity Aversion Literature: A Critical Assessment," *Economics and Philosophy*, 25, pp. 249-284, 2009.



- [2] Aumann, Robert and Michael Maschler and Richard Stearns, Repeated Games with Incomplete Information, MIT Press, Cambridge, 1995.
- [3] Bewley, Truman. “Knightian Decision Theory. Part I,” *Decisions in Economics and Finance*, 25, pp. 79-110, 2002.
- [4] Bodoh-Creed, Aaron. “Ambiguous Beliefs And Mechanism Design,” mimeo, Cornell University, 2011.
- [5] Bose, Subir and Emre Ozdenoren and Andreas Puppe. “Optimal Auctions with Ambiguity,” *Theoretical Economics*, 1, pp. 411-438, 2006.
- [6] Bose, Subir and Arup Daripa. “A Dynamic Mechanism And Surplus Extraction Under Ambiguity,” *Journal of Economic Theory*, 144, pp. 2084-2114, 2009.
- [7] Di Tillio, Alfredo and Nenad Kos and Matthias Messner. “The Design of Ambiguous Mechanisms,” work in progress, 2011.
- [8] Forges, Françoise. “An Approach to Communication Equilibria,” *Econometrica*, 54, pp. 1375-1385, 1986.
- [9] Forges, Françoise. “Universal Mechanisms,” *Econometrica*, 58, pp. 1341-1364, 1990.
- [10] Gilboa, Itzhak and David Schmeidler. “Maxmin Expected Utility with a Non-Unique Prior,” *Journal of Mathematical Economics*, 18, pp. 141-153, 1989.
- [11] Gilboa, Itzhak and David Schmeidler. “Updating Ambiguous Beliefs,” *Journal of Economic Theory*, 59, pp. 33-49, 1993.
- [12] Hanany, Eran and Peter Klibanoff, “Updating Preferences With Multiple Priors,” *Theoretical Economics*, 2, pp. 261-298, 2007.
- [13] Lopomo, Giuseppe and Luca Rigotti and Chris Shannon. “Uncertainty in Mechanism Design,” mimeo, 2010.

- [14] Marinacci, Massimo and Fabio Maccheroni and Aldo Rustichini. "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, pp. 1447-1498, 2006.
- [15] Myerson, Roger. "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47, pp. 61-73, 1979.
- [16] Myerson, Roger. "Optimal Auction Design," *Mathematics of Operations Research*, 6, pp. 58-73, 1981
- [17] Myerson, Roger. "Multistage Games with Communication," *Econometrica*, 54, pp. 323-358, 1986.
- [18] Pires, Cesaltina Pacheco. "A Rule For Updating Ambiguous Beliefs," *Theory and Decision*, 53, pp. 137-152, 2002.
- [19] Salo, Ahtia and Martin Weber. "Ambiguity Aversion in First Price Sealed Bid Auctions," *Journal of Risk and Uncertainty*, 11, pp. 123-137, 1995.
- [20] Saran, Rene. "Menu-Dependent Preferences and Revelation Principle," *Journal of Economic Theory*, forthcoming.
- [21] Siniscalchi, Marciano. "Two Out Of Three Ain't Bad: A Comment On "The Ambiguity Aversion Literature: A Critical Assessment,"" *Economics and Philosophy*, 25, pp. 335-356, 2009.
- [22] Siniscalchi, Marciano. "Dynamic Choice under Ambiguity," *Theoretical Economics*, forthcoming.