J. Econ. Theory 216 (2024) 105782

Contents lists available at ScienceDirect

Journal of Economic Theory

journal homepage: www.elsevier.com/locate/jet

Mechanism design with belief-dependent preferences ☆

Ernesto Rivera Mora

Cowles Foundation for Research in Economics, P.O. Box 208281, New Haven, CT 06520, United States of America

ARTICLE INFO

JEL classification:
D82
D91

Keywords:
Psychological game theory
Belief-dependent preferences
Revelation principle
Auctions with image concerns
Mechanism design with after-games

ABSTRACT

This paper studies mechanism design when agents have belief-dependent preferences, in that utilities depend on the agents' hierarchical posterior beliefs about types. For instance, agents may be subject to temptation, shame, image concerns, or privacy concerns. In this setting, the textbook revelation principle does not hold, since mechanisms can provide agents with information that affects posterior beliefs. This paper uses a psychological game framework suited for mechanism design, and provides a novel version of the revelation principle for belief-dependent preferences. The new revelation principle makes use of extended direct mechanisms that map each reported type into material outcomes and private suggestions of what posterior beliefs the agents should have. The paper shows that it suffices to use extended direct mechanisms that satisfy three conditions: *Bayesian incentive compatibility, individual rationality*, and a new condition called *believability*. The new revelation principle is used to find revenue-maximizing auctions when bidders have different types of image concerns. Moreover, it provides an alternate tool—distinct from Myerson's communication revelation principle—to study mechanism design with after-games.

1. Introduction

Psychological game theory has a long history, going back to the seminal work of Geanakoplos, Pearce, and Stacchetti (1989) whose framework was significantly expanded by Battigalli and Dufwenberg (2009). According to the later, for instance, an agent with image concerns may desire to be perceived by others as altruistic, healthy, wise, etc. Likewise, an agent with privacy concerns may be averse to revealing his level of income or consumption patterns. In these examples, the agent's preferences depend on his beliefs about what others believe about him.

These belief-dependent motivations have important implications for mechanism design. For instance, image concerns can induce pro-social behavior (Bénabou and Tirole, 2006). In turn, this can lead to increased donations to social projects when the donors' identities are observed (Alpizar, Carlsson, and Johansson-Stenman, 2008), or increased bidding in charity and art auctions when the bidders' identities are revealed (Bos and Pollrich, 2022). Privacy concerns can lead individuals to evade questions regarding sensitive personal information (Warner, 1965), or to protect private information that can be exploited for financial gain (Pai and Roth, 2013; Dziuda and Gradwohl, 2015).

Despite their importance, belief-dependent motivations are not accounted for in the standard mechanism design framework. A standard assumption in the literature is that agents' preferences depend on (payoff) types and material outcomes. Preferences

☆ I thank the Editor (Pierpaolo Battigalli), the Associate Editor, and the two referees. In addition, I thank Amanda Friedenberg, Andreas Blume, Inga Deimen, Martin Dufwenberg, Elliot Lipnowski, Philipp Strack, Rachel Mannahan, and Ambika Athreya for all their useful comments and suggestions.

E-mail address: ernesto.riveramora@yale.edu.

https://doi.org/10.1016/j.jet.2023.105782
Received 21 November 2021; final version received 18 December 2023; Accepted 19 December 2023

Available online 9 January 2024

depend on beliefs only insofar as beliefs impact the expectation of types and material outcomes. Beliefs do not enter directly in the agents' Bernoulli utility function. For example, in standard auction settings, utility functions depend on agents' valuations of the auctioned object, whether or not they win the auction, and how much they must pay. These utility functions do not directly depend on the agents' beliefs about valuations or other traits. As a result, they do not model image concerns, privacy concerns, or other belief-dependent motivations.

This paper explores the implications of belief-dependent preferences for mechanism design—especially, for the revelation principle and the use of direct mechanisms. A direct mechanism requires agents to report their types to the designer, and the designer to allocate material outcomes based on the reports. One implication of the textbook revelation principle is that, in order to solve a mechanism design problem, it suffices to restrict attention to direct mechanisms and truth-telling.[1] However, the textbook revelation principle fails in environments with belief-dependent preferences. The reason is that (textbook) direct mechanisms tell agents nothing about the other agents' reports. While this fact is inconsequential in the standard setting, revealing information about types may impact payoffs and behavior in environments with belief-dependent preferences. For instance, a participant in an art auction who desires to be perceived as an "art enthusiast" would presumably bid differently in auctions where bids are public as compared to those where bids are private. This paper shows that the informational limitations of (textbook) direct mechanisms render them insufficient to "cover" all the outcomes of Bayesian equilibria in belief-dependent settings. Thus, the textbook revelation principle does not apply.

The main result of this paper establishes a novel version of the revelation principle for belief-dependent environments. It states that all equilibrium outcomes are generated by truth-telling in a class of *extended direct mechanisms*. In an extended direct mechanism, agents report their types and the designer both allocates the material outcomes and sends private messages to the agents. These private messages correspond to suggestions for the posterior beliefs that an agent should hold. In this way, the messages summarize the relevant information agents can acquire through some Bayesian equilibrium of some Bayesian game.

The result implies that, for partial implementation in belief-dependent environments, it suffices to focus on extended direct mechanisms that satisfy three conditions: *individual rationality*, *Bayesian incentive compatibility*, and a new condition called *believability*. Individual rationality and Bayesian incentive compatibility are the standard conditions that require agents to have incentives to participate and report their types truthfully. The believability condition requires that the extended direct mechanism satisfies an independence property. This condition can be viewed as an asymmetric-information version of the consistency condition found in Mathevet, Perego, and Taneva (2020), one that takes into account each agent's observable dimension of material outcomes. Believability implies that—as long as all agents tell the truth—the beliefs suggested by the direct mechanism coincide with the agents' posterior beliefs as derived via Bayes updating. That is, the condition implies that the messages sent by the mechanism are "believable."

As an application, the paper addresses an auction design problem. An auctioneer aims to design a revenue-maximizing auction. The auctioneer knows the nature of the bidders' belief-dependent preferences, but not their valuation of the object. For instance, consider an art auction in which the bidders have image concerns. The bidders care not only about the transfers they pay, and whether or not they receive the object, but also about the perceptions of fellow participants. Specifically, bidders wish to be perceived as if they had a high valuation of the piece of art, and are thus concerned about hierarchies of beliefs about valuations. A recent literature takes up this application. (See Bos and Truyts (2021); Bos and Pollrich (2022).) However, absent a revelation principle for psychological games, the literature was forced to study symmetric mechanisms and symmetric monotone equilibria.[2] The present paper provides a revenue characterization result for additively separable belief-dependent preferences, showing that the auctioneer is able to fully extract the psychological sub-utility from the bidders. Moreover, it points to revenue-maximizing auctions for different types of image concerns. For some types of image concerns, the optimal allocation rule mirrors that of the standard setting. However, the optimal way to reveal information is sensitive to the type of image concern: the decision to publicly reveal or hide the agents' bids hinges on the specific type of image concern that the agents may hold.

Proving the revelation principle requires studying Bayesian equilibria in dynamic psychological games, as in Battigalli and Dufwenberg (2009) and Battigalli, Corrao, and Dufwenberg (2019a). In a psychological game, the agents' utilities depend on endogenous beliefs—i.e., beliefs that are shaped by behavior. In the context of this paper, each mechanism induces a dynamic psychological game where (1) the agents' behavior influences their beliefs about types, and (2) the agents' utilities at terminal nodes depend on those beliefs. Notice, the paper restricts the utility functions to depend directly on beliefs about types, and not on beliefs about strategies. This assumption restricts the nature of the belief-dependent motivations—it cannot capture reciprocity, guilt, frustration, etc.

At the same time, the assumption that preferences depend on beliefs about types allows the paper to provide a new revelation principle for environments with after-games. In such environments, the agents play an after-game—or a sequence of after-games—following the mechanism. Examples include a monopolist expecting the buyer to resell the good (Calzolari and Pavan, 2006a), sequential contracting with multiple principals (Calzolari and Pavan, 2006b, 2009), a monopolist selling conspicuous goods that signal social status (Rayo, 2013), a consumer purchasing a durable good from a seller with limited commitment (Doval and Skreta, 2023), and bidders participating in an aftermarket following the auction (Giovannoni and Makris, 2014; Dworczak, 2020). In these settings, different posterior beliefs drive different outcomes in the after-game, so the information acquired or revealed by the mecha-

[1] By the "textbook revelation principle" I mean a set of results along the lines used in Myerson (1979), Gibbard (1973), and Dasgupta, Hammond, and Maskin (1979).

[2] Importantly, Bos and Pollrich (2022) show that symmetric monotone equilibria may not exist for some auctions.

nism becomes relevant. Notice that higher-order beliefs are relevant for studying after-games. This arises when equilibrium behavior in the after-game depends on the full hierarchy of beliefs, as in the global games literature (Carlsson and Van Damme, 1993), and information-sharing mechanisms (Rivera Mora, 2023).

Myerson (1982) provides a *communication revelation principle*, which applies to settings with after-games. There, a direct mechanism not only chooses a material outcome but also suggests actions that agents should play in the after-game. The communication revelation principle requires that such a mechanism satisfies individual rationality, Bayesian incentive compatibility, and obedience constraints. That is, agents must have incentives to participate, to report their types truthfully, and to obey the mechanism's recommendation. The revelation principle in this paper is similar, to the extent that it also provides suggestions. However, here the suggestions are about the beliefs the agents should have rather than the actions the agents should take. Thus, instead of obedience constraints, the mechanism must satisfy the believability condition. (Again, believability ensures that, in equilibrium, the suggested hierarchies of beliefs coincide with actual hierarchies.)

One might conjecture that all environments with belief-dependent preferences can be converted to environments with after-games and, if so, that the communication revelation principle applies. However, this conjecture is incorrect. Section 7.1 shows that the conjecture fails if i's belief-dependent utility function is not convex in i's hierarchies of beliefs (holding other agents' hierarchies of beliefs fixed). Many belief-dependent motivations have non-convex utility representations—e.g., temptation (Gul and Pesendorfer, 2001), ego utility (Köszegi, 2006), shame (Dillenberger and Sadowski, 2012), and some types of image concerns (Example 2.5). The key is that, in environments with after-games, more information cannot make an agent i worse off unless other players know i has more information. However, with a non-convex belief-dependent utility function, agents exhibit some information aversion.

In addition, the communication revelation principle cannot be applied to all environments with after-games. Applying it requires complete knowledge of the after-game. Yet, there are many environments where the agents do not know the after-game; in those cases, the agents may use heuristics to determine preferences about information. For instance, a consumer may not know the nature of future interactions with a retailer, but may nevertheless be averse to revealing her valuation of the retailer's products. Similarly, a consumer of a conspicuous good may not know the nature of future social interactions, but may nevertheless want to be perceived as belonging to a high social stratum (Rayo, 2013). Moreover, even when the communication revelation principle can be used, it may be computationally useful to consider a reduced-form approach to model the outcomes of the after-game. The martingale properties of posterior hierarchies of beliefs make them convenient for solving optimization problems. For instance, Dworczak (2020) uses a reduced-form approach to solve for optimal auctions with resale opportunities, while Rivera Mora (2023) uses a reduced-form approach to characterize environments where mechanisms help agents to share information.

This paper fits into a broader recent literature that analyzes environments not covered by different versions of the revelation principle. See Saran (2011), Doval and Skreta (2022), Lipnowski and Mathevet (2018), Mathevet, Perego, and Taneva (2020), Crawford (2021), and Sugaya and Wolitzky (2021). The paper closest to this one is Doval and Skreta (2022). As in their paper, the revelation principle provided here is based on posterior beliefs. However, there are important differences. Their revelation principle applies to a setting where (1) there is a principal and an agent with private information, (2) reduced-form preferences are generated by a continuation mechanism design environment, and (3) all signals are public. In contrast, this paper allows for settings with (1) multiple agents, each with private information, (2) general belief-dependent preferences, and (3) both private and public signals. So, the revelation principle here is suitable for the analysis of belief-dependent frameworks which have not been covered in the literature.

2. Application: auctions with image concerns

An auctioneer offers an indivisible piece of artwork to n agents. Each agent has a private valuation of the piece of art. The agents have image concerns: they care about other participants' perceptions of how they value the artwork. So, overall, they care about receiving the object, their payments, and about the other participants' perception of their private valuation.

Let I denote the set of agents. For each agent $i \in I$, there is a finite set of possible non-negative valuations denoted by Θ_i. The realized valuation of i is independently drawn from a full-support type distribution $\overline{\mu}_i \in \Delta(\Theta_i)$ and is only known by i. It is transparent that valuations are distributed according to the product measure $\mu = \bigotimes_{i \in I} \overline{\mu}_i$.

The utility function of agent i is given by

$$u_i(\theta_i, x_i, t_i, h_i) = \theta_i \cdot x_i - t_i + f_i(h_i),$$

where θ_i is i's valuation of the object, $x_i \in \{1, 0\}$ is an indicator variable that equals 1 if i gets the object, t_i stands for the transfer paid to the auctioneer, and h_i is i's posterior hierarchy of beliefs of valuations. The hierarchical posterior beliefs h_i are computed at the end of the mechanism by Bayesian updating. The term $\theta_i \cdot x_i - t_i$ represents i's material sub-utility and the term $f_i(h_i)$ represents i's psychological sub-utility. The psychological sub-utility can capture i's image concerns. Five examples are introduced: one where i does not have belief-dependent preferences and four with different forms of image concerns.

Example 2.1. Standard preferences: Here, there are no belief-dependent preferences. So, f_i is the zero mapping. This case serves as a reference point for analyzing how classical results change when image concerns are present.

Example 2.2. Simple image concerns: Regardless of i's true valuation, i wants to be perceived as having a valuation of the artwork above some benchmark $b > 0$. So, i receives a psychological reward if i's expectation of j's expectation of i's valuation is above the benchmark. The psychological sub-utility of i is given by

$$f_i(h_i) = a \sum_{j \in I, j \neq i} \mathbb{1}\left[\mathbb{E}_i\left[\mathbb{E}_j[\theta_i] \mid h_i\right] \geq b\right],$$

where $a > 0$ is the psychological reward and $\mathbb{E}_i\left[\mathbb{E}_j[\theta_i] \mid h_i\right]$ denotes i's expectation of j's expectation of θ_i when i has hierarchy h_i. Notice that f_i only depends on i's second-order beliefs.

Example 2.3. Expectation-based image concerns: Regardless of i's true valuation, agent i wants to be perceived as having a higher valuation of the artwork. This perception is captured by i's expectation of j's expectation of i's valuation. The psychological sub-utility of i is given by

$$f_i(h_i) = a \sum_{j \in I, j \neq i} \mathbb{E}_i\left[\mathbb{E}_j[\theta_i] \mid h_i\right],$$

where $a > 0$ indicates the intensity of the image concerns. Notice, similarly to simple image concerns, f_i only depends on i's second-order beliefs. However, it differs from simple image concerns in that f_i is strictly increasing and continuous in i's second-order expectation.

Example 2.4. Sophisticated-type image concerns:[3] Agent i seeks to be perceived as a sophisticated art aficionado, i.e., as someone who genuinely values art. The agent is perceived as sophisticated if it is commonly believed that i's value is above some benchmark $b > 0$. The psychological sub-utility of i is given by

$$f_i(h_i) = \begin{cases} a & \text{if } i \text{ is certain that there is common belief that } \theta_i \geq b \\ 0 & \text{otherwise,} \end{cases}$$

where $a > 0$ is the reward of being perceived as sophisticated. Notice that this differs from simple image concerns. Since the event $\theta_i \geq b$ has to be commonly believed, f_i depends on the entire hierarchy of beliefs (not just second-order beliefs).

Example 2.5. Unsophisticated-type image concerns: Agent i does not want to be perceived as unsophisticated, i.e., as someone who does not truly appreciate art. The agent is perceived as unsophisticated if there is common belief that i's value is below some benchmark $b > 0$. The psychological sub-utility of i is given by

$$f_i(h_i) = \begin{cases} 0 & \text{if } i \text{ is certain that there is common belief that } \theta_i < b \\ a & \text{otherwise,} \end{cases}$$

where $a > 0$ is the reward of not being perceived as unsophisticated. Similar to sophisticated-type image concerns, f_i depends on i's entire hierarchy of beliefs. The key distinction is that, here, uncertainty regarding valuations is favorable, whereas with sophisticated-type image concerns, uncertainty is unfavorable.

The auctioneer wishes to find a mechanism that maximizes her expected revenue. A mechanism is a sequential game with chance moves. In the mechanism, a material outcome for i, $(x_i, t_i) \in Y_i$, describes whether agent i acquires the object ($x_i = 1$) or not ($x_i = 0$) and the transfer $t_i \in \mathbb{R}$ that i pays to the auctioneer. The set of collective material outcomes is $Y := \left\{(x_i, t_i)_{i \in I} \in \prod_{i \in I} Y_i : \sum_{i \in I} x_i \leq 1\right\}$; this restricts the object to be allocated to at most one agent.[4] Agent i necessarily observes his own material outcome $(x_i, t_i) \in Y_i$, but may not observe $(x_j, t_j) \in Y_j$ for $j \neq i$. Whether or not the agent has any information depends on the specific mechanism. The auctioneer's revenue from a collective material outcome $(x_i, t_i)_{i \in I} \in Y$ is $\sum_{i \in I} t_i$. The auctioneer only cares about maximizing her revenue.

All agents have an exogenous outside option with zero value. The agents' psychological sub-utility is motivated by the social component of the interaction. So, the agents receive their psychological sub-utility only if they are physically present in the auction and are observed by the other bidders. If they decide to not participate, they forfeit both their material and psychological sub-utility, leading to zero utility. A Bayesian equilibrium of a mechanism is individually rational if it incentivizes agents to participate, i.e., if each type has non-negative expected utility.[5] Therefore, the auctioneer's objective is to find a mechanism and an associated individually rational Bayesian equilibrium that maximizes her expected revenue.

2.1. Failure of the revelation principle

In the standard framework with partial implementation—where preferences are belief-independent and the auctioneer freely chooses the equilibrium—there is no loss in focusing on direct mechanisms. A direct mechanism is a mapping $\mathcal{M} : \Theta \to \Delta(Y)$ that corresponds to a game form where:

[3] The use of the term "sophisticated" here differs from its use in the context of behavioral economics, where it often describes an individual with self-awareness of their own behavioral biases.

[4] Notice, the mechanism has chance moves. This is consistent with stochastic mechanisms.

[5] A discussion for using Bayesian equilibrium as a solution concept can be found in Section 7.4.

(i) Agents privately and simultaneously report their valuations to the mechanism.
(ii) The material outcome is drawn according to the distribution specified by $\mathcal{M}$ given the reported values.

The textbook revelation principle states that it is without loss of generality to analyze direct mechanisms in which agents have an incentive to participate and truthfully report their valuations.

Direct mechanisms do not explicitly describe what information is revealed to the agents after the material outcome is selected. In particular, it is not specified whether agent i observes the other agents' reported values. In the standard framework, this omission has no bearing, as changes in terminal information are inconsequential for utility and behavior. However, in settings with belief-dependent preferences, this omission does have consequences. As highlighted by Battigalli and Dufwenberg (2009, 2022), utility and behavior in psychological games are sensitive to terminal information. Therefore, for accurately characterizing equilibria, it is necessary to describe the terminal information.

The designer has multiple ways to specify the agents' terminal information of direct mechanisms. One natural alternative is to conceal all information, precluding each agent i from observing the reported type profile θ_{-i} or the material outcome y_{-i} of the other agents. A second natural alternative is to reveal all information, allowing agents to directly observe the reported type profile θ and the collective material outcome y. Under belief-dependent preferences, these two extreme descriptions of terminal information are insufficient to account for all Bayesian equilibria. There are Bayesian equilibria of some mechanisms where only partial information is revealed. As a result, the textbook revelation principle fails if the designer uses direct mechanisms that only embody extreme descriptions of terminal information. The example below illustrates this point in an auction setting with image concerns.

Example 2.6. There are two agents: Ann (A) and Bob (B). Ann's set of valuations is $\Theta_A = \{0, 6\}$ while Bob's is $\Theta_B = \{0\}$. The type distribution $\overline{\mu}_A \in \Delta(\Theta_A)$ assigns equal probability to each valuation of Ann. Ann's set of material outcomes is $Y_A = \{(x_A, t_A) : (x_A, t_A) \in \{0, 1\} \times \mathbb{R}\}$ while Bob's set of material outcomes is $Y_B = \{(0, 0)\}$; so Bob is just a spectator. Only Ann has belief-dependent preferences, which are given by simple image concerns (Example 2.2) with a reward of $a = 4$ and a benchmark of $b = 4$. So, Ann has a psychological sub-utility of 4 if and only if her expectation of Bob's expectation of her valuation equals or exceeds 4.

Consider the following mechanism:

> **Mechanism** $(*)$**.** Ann privately reports her valuation to the auctioneer. If Ann reports a valuation of 0, then Ann pays 2 and no agent receives the object. If Ann reports 6, then Ann pays 10 and Ann gets the object. Bob does not observe Ann's report, Ann's payment, or whether Ann receives the object. At the end of the mechanism, the auctioneer sends a public message to both agents. If Ann reported 6, the auctioneer sends a message H. If Ann reported 0, then the auctioneer sends a message H or L with equal probability.

Suppose that Ann truthfully reports her valuation in mechanism $(*)$. After observing message H (resp. message L), Ann has an expected value 4 (resp. an expected value of 0) of Bob's expectation of Ann's valuation. Hence, Ann receives a psychological sub-utility of 4 whenever the message H is sent and 0 otherwise. As a consequence, the honest strategy is a Bayesian equilibrium in which the auctioneer has an expected payoff of 6, Bob has an expected utility of 0, and each of Ann's types has an expected utility of 0.

Lemmata B.1 and B.2 show that the honest equilibrium of mechanism $(*)$ is not represented by direct mechanisms that fully disclose or fully conceal information. To generate the expected utilities corresponding to such an equilibrium, an ex-ante expected psychological sub-utility of 3 must be ensured for Ann. However, direct mechanisms that fully reveal (resp. fully conceal) reports generate expected psychological sub-utility of 2 (resp. sub-utility of 0) for Ann. Therefore, this class of mechanisms fail to capture the equilibrium described above.

The key to formulating a revelation principle for belief-dependent preferences lies in using mechanisms that account for any intermediate amount of information. Towards this end, Section 4 introduces a class of "extended" direct mechanisms which not only choose the material outcome but also send private messages that convey explicit information. Theorem 5.1 shows that this class covers all equilibria of all mechanisms, thereby establishing a revelation principle.

2.2. Revenue maximization

The implications of Example 2.6 extend beyond the revelation principle, impacting revenue maximization as well. In that example, direct mechanisms that either fully disclose or fully conceal information are insufficient for maximizing revenue. The maximum expected revenue can only be achieved when the mechanism partially reveals Ann's type. (See Lemma 6.1.)

By applying the revelation principle, Section 6 offers a revenue characterization result for settings with belief-dependent preferences. This characterization provides criteria for maximizing revenue, describing not only the optimal way to allocate the object but also the optimal way to reveal information.

For instance, in settings with expectation-based, sophisticated-type, or unsophisticated-type image concerns, the optimal allocation rule mirrors that of the standard setting. However, the optimal way to reveal information depends on the specific type of image concerns. In settings with expectation-based image concerns, the information revealed does not impact revenue. Hence, the disclosure of agents' bids is inconsequential. In contrast, in settings with sophisticated-type and unsophisticated-type image concerns, the information revealed by the mechanism does impact revenue. In particular, the choice of what information is revealed (e.g., revealing the bids) plays a pivotal role for revenue maximization.

3. Model

Throughout the paper, take the following conventions. Endow a compact metric space C with its Borel sigma algebra. Denote by $\Delta(C)$ the set of probability measures on C and endow $\Delta(C)$ with the topology of weak convergence. Endow the product of topological spaces with the product topology. In addition, for any given set of indexes I and family of sets $(B_i)_{i\in I}$, write $B_{-i} := \prod_{j\in I\setminus\{i\}} B_j$.

3.1. Environment

There is a finite set of agents I. Each agent i cares about material outcomes and the profile of types. For each agent i there is a set of agent-specific material outcomes Y_i. The set of collective material outcomes is $Y \subseteq \prod_{i\in I} Y_i$.[6] Each agent i has a finite set of types Θ_i. Write $\Theta := \prod_{i\in I} \Theta_i$ for the set of type profiles. The realized type profile is drawn from a full support common prior $\mu \in \Delta(\Theta)$.

In addition, agents also care about their hierarchy of beliefs about Θ. The first-order belief h_i^1 describes the probability that agent i assigns to types in Θ_{-i}; the second order belief h_i^2 describes the probability that agent i assigns to both the types in Θ_{-i} and $-i$'s first-order beliefs; and so on for higher-order beliefs. Write H_i for i's **set of collectively coherent hierarchies of beliefs** and call $H := \prod_{i\in I} H_i$ the **belief structure**. Write $h_i = (h_i^1, h_i^2, ...) \in H_i$ for a **hierarchy of beliefs** of agent i and call h_i^k the k^{th}**-order belief** of i. Appendix A.1 constructs the belief structure.

Ex-ante, the hierarchies of beliefs are induced by the common prior $\mu \in \Delta(\Theta)$. However, the agents' interaction may reveal information about the agents' moves and that information can impact the agents' posterior hierarchies of beliefs. These posterior beliefs will be determined in equilibrium.

Each agent i has an utility function $u_i : \Theta \times Y \times H_i \to \mathbb{R}$ where H_i captures the set of i's posterior hierarchies of beliefs, i.e., after the agents' interaction takes place. In addition, each type θ_i has an outside option with value $\overline{u}_i(\theta_i) \in \mathbb{R} \cup \{-\infty\}$. The case $\overline{u}_i(\theta_i) = -\infty$ has the interpretation that there is no outside option for type θ_i. Finally, the mechanism designer has a payoff function $\pi : \Theta \times Y \times H \to \mathbb{R}$, where H captures the posterior profile of hierarchies of beliefs.

3.2. The extensive form

The designer chooses a mechanism, i.e., a set of rules that, together with the agents' behavior, determines a collective material outcome $y \in Y$ and each agent i's ex-post information. The mechanism is represented by an extensive form Γ with perfect recall and (potentially) chance moves. For the purposes of this paper, the full definition of the extensive form will not be needed. Instead, there will be two necessary ingredients: the derived strategies and the terminal information sets.[7]

For each agent i, let S_i be a finite set of pure strategies for i in Γ. Let S_0 be a finite set of pure strategies for chance. There is an exogenous distribution $p_0 \in \Delta(S_0)$ that describes chance's behavior in Γ. Write $S := \prod_{i\in I} S_i$. The set $S_0 \times S$ is the set of strategy profiles.

The idea is as follows: When agents interact in the mechanism, they can be thought of as choosing a strategy in the extensive form Γ. Likewise, chance interacts by selecting a strategy based on the distribution p_0. The realized strategy profile determines a terminal node which, in turn, leads to a collective material outcome $y \in Y$ and a terminal information set for each agent i.

With this in mind, let Z be the set of terminal nodes. Let $\zeta : S_0 \times S \to Z$ be the **path function**, mapping each strategy profile (s_0, s) to the terminal node $\zeta(s_0, s)$ that it reaches. Let $\gamma : Z \to Y$ describe the **material outcome function**, mapping each terminal node into the induced collective material outcome. Similarly, let $\gamma_i : Z \to Y_i$ be the **material outcome function for** i.

For each agent i, let $\mathcal{P}_i$ denote the partition of Z representing the collection of **terminal information sets** of agent i. Each partition $\mathcal{P}_i$ must ensure that the outcome function $\gamma_i : Z \to Y_i$ is $\mathcal{P}_i$-measurable. That is, all the terminal nodes in a partition element of $\mathcal{P}_i$ are associated with the same material outcome $y_i \in Y_i$. So, in this sense, at the end of the play, each agent i observes his material outcome y_i. However, depending on i's partition, i may or may not observe the material outcomes of other agents. Notice, in the standard setting, information about Y_{-i} is irrelevant. However, with belief-dependent preferences, information about Y_{-i} becomes relevant as it conveys information about the agents' actions and, indirectly, information about the agents' types.

3.3. The Bayesian game

The extensive form Γ induces a game of incomplete information. In this game, Nature determines the type profile $\theta \in \Theta$. Each agent i observes θ_i but not θ_{-i}. Agents then play the game given by the extensive form Γ. The set $\Theta \times Z$ represents the terminal nodes in the tree representation of the induced game of incomplete information. The types Θ and the partition $\mathcal{P}_i$ define a partition $\hat{\mathcal{P}}_i$ on $\Theta \times Z$ with partition members $\hat{\rho}_i = \{\theta_i\} \times \Theta_{-i} \times \rho_i$ for each $(\theta_i, \rho_i) \in \Theta_i \times \mathcal{P}_i$. So, if i is of type θ_i and the terminal node $\zeta(s_0, s)$ is reached, i observes his type θ_i and the terminal information set $\rho_i \in \mathcal{P}_i$ that contains $\zeta(s_0, s)$. Because elements of $\hat{\mathcal{P}}_i$ are associated

[6] Letting Y be a strict subset of $\prod_{i\in I} Y_i$ allows for a rich set of applications. See Discussion 7.2.

[7] Each sequential game represented in extensive form is associated with information that is revealed to the agents. This information can be revealed either during the course of the interaction or after the interaction concludes. For instance, in an English auction, the agents observe the bidding process during the course of the auction. In a sealed-bid auction, the agents observe the bids after the auction concludes (provided the auctioneer chooses to reveal the bids). By assuming that the information structure in the extensive-form representation satisfies perfect recall, the paper implicitly assumes that the agents remember all information provided during the course of the interaction. (See Battigalli and Generoso (2021).) Thus, the terminal information sets capture both the information provided during the course of interaction and the information provided after the interaction concludes.

with elements of $\mathcal{P}_i$, refer to elements of $\hat{\mathcal{P}}_i$ as terminal information sets of the game of incomplete information. Write $\hat{\mathcal{P}} := \prod_{i\in I} \hat{\mathcal{P}}_i$. For each $\hat{\rho} = (\hat{\rho}_i)_{i\in I} \in \hat{\mathcal{P}}$, write $\bigcap \hat{\rho}$ for the set $\bigcap_{i\in I} \hat{\rho}_i$, i.e., the set of terminal nodes consistent with the tuple $(\hat{\rho}_i)_{i\in I}$.

The common prior μ induces a type structure $(\Theta_i, \beta_i)_{i\in I}$, where each $\beta_i : \Theta_i \to \Delta(\Theta_{-i})$ is a belief map for agent i so that, for each $\theta_i \in \Theta_i$,

$$\beta_i(\theta_i)(\theta_{-i}) := \frac{\mu(\theta_i, \theta_{-i})}{\operatorname{marg}_{\Theta_i} \mu(\theta_i)}.$$

The game of incomplete information induced by Γ and the type structure $(\Theta_i, \beta_i)_{i\in I}$ define a **Bayesian game.** With a slight abuse of notation, write (Γ, μ) for this induced Bayesian game. Within the Bayesian game, a **behavior strategy** of agent i is a mapping $\sigma_i : \Theta_i \to \Delta(S_i)$.[8]

It will be convenient to describe how the elements of $\hat{\mathcal{P}}$ are associated with types and material outcomes. To do so, write $\operatorname{proj}_{\Theta_i}(\hat{\rho}) = \theta_i$ for the unique type $\theta_i \in \Theta_i$ that satisfies $\hat{\rho}_i \subseteq \{\theta_i\} \times \Theta_{-i} \times Z$. Similarly, write $\operatorname{proj}_{Y_i}(\hat{\rho}) = y_i$ for the unique material outcome $y_i \in Y_i$ so that $(\theta, z) \in \hat{\rho}_i$ implies $\gamma_i(z) = y_i$. (Recall that γ_i is $\mathcal{P}_i$-measurable.) Write $\operatorname{proj}_{\Theta}(\hat{\rho}) := (\operatorname{proj}_{\Theta_i}(\hat{\rho}))_{i\in I}$ and $\operatorname{proj}_{Y}(\hat{\rho}) := (\operatorname{proj}_{Y_i}(\hat{\rho}))_{i\in I}$. Note, since the set $\hat{\mathcal{P}}$ is finite, the mappings $\operatorname{proj}_{\Theta} : \hat{\mathcal{P}} \to \Theta$, $\operatorname{proj}_{\Theta_i} : \hat{\mathcal{P}} \to \Theta_i$, $\operatorname{proj}_{Y} : \hat{\mathcal{P}} \to Y$, and $\operatorname{proj}_{Y_i} : \hat{\mathcal{P}} \to Y_i$ are measurable.

3.4. Bayesian equilibrium

This subsection defines Bayesian equilibrium. In the standard framework, the definition of Bayesian equilibrium builds on two ingredients: strategies and interim beliefs. Here the definition also builds on terminal beliefs—mappings that specify the beliefs that agents would have given the information learned at the end of the game. Bayesian equilibrium imposes two requirements on strategies and beliefs. First, the agents must optimize given the terminal belief mappings and the others' strategy profile; this optimization depends on the type profiles, the material outcomes, and the hierarchies of beliefs at the end of the Bayesian game. Second, the agents' belief mappings must be consistent with the strategies played.

A **terminal belief mapping** for agent i is a function $\hat{\beta}_i : \hat{\mathcal{P}}_i \to \Delta(\hat{\mathcal{P}}_{-i})$ so that for each $\hat{\rho}_i \in \hat{\mathcal{P}}_i$,

$$\hat{\beta}_i(\hat{\rho}_i)\left(\left\{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i} : \bigcap(\hat{\rho}_i, \hat{\rho}_{-i}) \neq \emptyset\right\}\right) = 1.$$

That is, $\hat{\beta}_i(\hat{\rho}_i)$ assigns positive probability only to the terminal information sets in $\hat{\mathcal{P}}_{-i}$ that are consistent with $\hat{\rho}_i$. The terminal belief mappings $\hat{\beta} = (\hat{\beta}_i)_{i\in I}$ induce **terminal hierarchies of beliefs**, i.e., a function $\hat{\delta}_i : \hat{\mathcal{P}}_i \to H_i$ which maps each terminal information set to the agents' posterior hierarchies of beliefs induced by $\hat{\beta}$. Appendix A shows how $\hat{\delta} = (\hat{\delta}_i)_{i\in I}$ is derived from $\hat{\beta}$. Notice, given a hierarchy mapping $\hat{\delta}_i$, i's utility at each $\hat{\rho} \in \hat{\mathcal{P}}$ is $u_i(\operatorname{proj}_{\Theta}(\hat{\rho}), \operatorname{proj}_{Y}(\hat{\rho}), \hat{\delta}_i(\hat{\rho}_i))$.

Fix a profile of behavior strategies of the co-players σ_{-i} and terminal beliefs $\hat{\beta}$. Notice, when a type θ_i chooses his strategy s_i, θ_i is effectively choosing a distribution over the information sets $\hat{\rho} \in \hat{\mathcal{P}}$ that satisfy $\operatorname{proj}_{\Theta_i}(\hat{\rho}) = \theta_i$. With this in mind, write $\mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta})$ for the interim expected utility of type θ_i when θ_i plays s_i, other agents play σ_{-i}, and the terminal beliefs are determined by $\hat{\beta}$. (See Appendix A.2 for its calculation.) With a slight abuse of notation, write $\mathbb{E}u_i(\theta_i, \sigma_i \mid \sigma_{-i}, \hat{\beta})$ for $\sum_{s_i \in S_i} \mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta}) \cdot \sigma_i(\theta_i)(s_i)$.

Note that, in principle, the terminal beliefs $\hat{\beta}$ may not be consistent with the strategy profile σ played. The concept of Bayesian equilibrium requires that they are. To formally define this, write $\mathbb{P}(\cdot \mid s_i, \sigma_{-i}) \in \Delta(\hat{\mathcal{P}})$ for the ex-ante probability measure on $\hat{\mathcal{P}}$, given that i plays s_i and co-players choose σ_{-i}. So, $\mathbb{P}(\hat{\rho} \mid s_i, \sigma_{-i})$ is the ex-ante probability of reaching all the terminal information sets $\hat{\rho} = (\hat{\rho}_i)_{i\in I}$ given that agent i plays s_i and agents $-i$ play the behavior strategies σ_{-i}. (See Appendix A.2 for its calculation.)

Definition 3.1. The terminal beliefs $\hat{\beta} = (\hat{\beta}_i)_{i\in I}$ are **consistent** with the strategy profile σ if, for each $s_i \in S_i$,

$$\hat{\beta}_i(\hat{\rho}_i)(\hat{\rho}_{-i}) \cdot \mathbb{P}(\{\hat{\rho}_i\} \times \hat{\mathcal{P}}_{-i} \mid s_i, \sigma_{-i}) = \mathbb{P}((\hat{\rho}_i, \hat{\rho}_{-i}) \mid s_i, \sigma_{-i}).$$

Consistency requires that $\hat{\beta}_i(\hat{\rho}_i)(\hat{\rho}_{-i})$ is the conditional probability of $\hat{\rho}_{-i}$ given $\hat{\rho}_i$ under the probability measure $\mathbb{P}(\cdot \mid s_i, \sigma_{-i})$. Call the profile $(\hat{\beta}, \sigma)$ **consistent** if $\hat{\beta}$ is consistent with σ. Notice, consistency imposes the implicit requirement that beliefs satisfy *own-action independence:* The probability that i assigns to $\hat{\rho}_{-i}$ is independent of which strategy s_i is used (provided that i uses a strategy that allows $\hat{\rho}_i$). So, if i deviates from the strategy profile σ, i still believes that other agents are playing in accordance with σ_{-i}.

Definition 3.2. Call $(\sigma, \hat{\beta})$ a **Bayesian equilibrium** of the Bayesian game (Γ, μ) if

(i) for each $\theta_i \in \Theta_i$ and each strategy σ_i', $\mathbb{E}u_i(\theta_i, \sigma_i \mid \sigma_{-i}, \hat{\beta}) \geq \mathbb{E}u_i(\theta_i, \sigma_i' \mid \sigma_{-i}, \hat{\beta})$, and

(ii) the terminal beliefs $\hat{\beta}$ are consistent with σ.

[8] Since σ_i specifies a mixed strategy for every type, it does not correspond to a behavior strategy in the usual sense. However, if Γ has simultaneous moves, it corresponds to a behavior strategy in the usual sense.

It is important to remark that the notion of Bayesian equilibrium may be unsatisfactory for dynamic Bayesian games. Under Bayesian equilibrium, agents optimize at the root, but may not be optimizing at other non-terminal information sets. Nevertheless, the revelation principle applies to any dynamic refinement of Bayesian equilibrium. (See Discussion 7.4.)

3.5. The designer's optimization problem

A Bayesian equilibrium $(\sigma, \hat{\beta})$ of the Bayesian game (Γ, μ) satisfies **individual rationality** if, for each $\theta_i \in \Theta_i$, $\mathbb{E}u_i(\theta_i, \sigma_i \mid \sigma_{-i}, \hat{\beta}) \geq \overline{u}_i(\theta_i)$. That is, the expected utility of θ_i under $(\sigma, \hat{\beta})$ must be weakly higher than the outside option's value $\overline{u}_i(\theta_i)$.[9]

The designer can only choose Bayesian equilibria that satisfy individual rationality, otherwise some agent would prefer to not participate in the mechanism and instead take his outside option.

Write $\mathbb{E}\pi(\sigma, \hat{\beta})$ for the designer's expected utility of a Bayesian equilibrium $(\sigma, \hat{\beta})$. (See Appendix A.2 for its calculation.) The designer's goal is to find an extensive form Γ, and a profile $(\sigma, \hat{\beta})$ of the induced Bayesian game that maximizes $\mathbb{E}\pi(\sigma, \hat{\beta})$ subject to the constraint that $(\sigma, \hat{\beta})$ is a Bayesian equilibrium that satisfies individual rationality.

4. Extended direct mechanisms

An extended direct mechanism maps types to material outcomes and individual messages. The set of messages for agent i is a finite set $M_i \subseteq H_i$. A message is interpreted as a suggestion of the hierarchies of beliefs that agent i should hold.

Definition 4.1. An **extended direct mechanism** is a mapping $\mathcal{M} : \Theta \to \Delta(Y \times H)$ so that, for each $\theta \in \Theta$, $\mathcal{M}(\theta)$ has finite support.

Each extended mechanism $\mathcal{M}$ induces a set of **collective material outcomes** $\overline{Y} \subseteq Y$ and a set of **hierarchy-message profiles** $M \subseteq H$ given by

$$\overline{Y} := \bigcup_{\theta \in \Theta} \text{Supp}\,(\text{marg}_Y \mathcal{M}(\theta)) \quad \text{and} \quad M := \bigcup_{\theta \in \Theta} \text{Supp}\,(\text{marg}_H \mathcal{M}(\theta)).$$

Call $\overline{Y} \times M$ the **support** of $\mathcal{M}$. The set $\overline{Y}_i := \text{proj}_{Y_i} \overline{Y}$ represents the set of i's material outcomes that are feasible under $\mathcal{M}$, while $M_i := \text{proj}_{H_i} M$ corresponds to the set of i's hierarchy-messages that are feasible under $\mathcal{M}$. Since $\mathcal{M}(\theta)$ has finite support for each $\theta \in \Theta$, and Θ is finite, $\overline{Y} \times M$ is also finite.

Each extended direct mechanism $\mathcal{M}$ induces a **canonical extensive form** $\Gamma(\mathcal{M})$ described as follows: First, agent i reports his type, i.e., an element of Θ_i. The reports are private and simultaneous. Second, given the reported profile $\theta \in \Theta$, chance selects $(y, h) \in \overline{Y} \times M$ according to the distribution $\mathcal{M}(\theta)$. Third, the mechanism privately reveals the material outcome y_i and the hierarchy-message h_i to each agent i.

The canonical extensive form $\Gamma(\mathcal{M})$ has the following features: First, the set of pure strategies for agent i is $S_i = \Theta_i$. Second, $S_0 = (\overline{Y} \times M)^\Theta$ is the set of strategies of chance.[10] Third, the set of terminal nodes is $Z = \Theta \times \overline{Y} \times M$. Fourth, the distribution of chance's strategies p_0 is such that, given a report profile $s = \theta$, the probability that chance selects terminal node $(\theta, y, h) \in Z$ —which induces collective material outcome $y \in Y$—is $\mathcal{M}(\theta)(y, h)$. (See Lemma C.2 for the construction.) Finally, the partition $\mathcal{P}_i$ of terminal nodes consists of sets of the form $\{\theta_i\} \times \Theta_{-i} \times \{y_i\} \times \overline{Y}_{-i} \times \{h_i\} \times M_{-i}$. So, agent i observes (only) i's report θ_i, his material outcome y_i, and his message h_i. Note that the set of strategy profiles $S_0 \times S$ and terminal nodes Z are all finite.

Just as any extensive form, $\Gamma(\mathcal{M})$ induces a game of incomplete information. The information sets of agent i are given by a partition $\hat{\mathcal{P}}_i$ of the set $\Theta \times Z = \Theta \times (\Theta \times \overline{Y} \times M)$, where the first entry of Θ represents the set of types and the second entry of Θ represents the set of reports. Note that each terminal information set $\hat{\rho}_i \in \hat{\mathcal{P}}_i$ is of the form

$$\hat{\rho}_i = \{\theta_i\} \times \Theta_{-i} \times \{\theta_i'\} \times \Theta_{-i} \times \{y_i\} \times \overline{Y}_{-i} \times \{h_i\} \times M_{-i}.$$

So, in this sense, agent i observes his type θ_i, his report θ_i', his material outcome y_i, and his message h_i. For convenience, write $[\theta_i, \theta_i', y_i, h_i]$ for such a terminal information set $\hat{\rho}_i$ and write $[\theta, \theta', y, h] := ([\theta_i, \theta_i', y_i, h_i])_{i \in I}$.

Each extended direct mechanism $\mathcal{M}$ induces a **canonical Bayesian game** $(\Gamma(\mathcal{M}), \mu)$. In this Bayesian game, a strategy σ_i^* is **honest** if, for each $\theta_i \in \Theta_i$, $\sigma_i^*(\theta_i)(\theta_i) = 1$. So σ_i^* is honest if i reports his true type with probability one. Call $\sigma^* = (\sigma_i^*)_{i \in I}$ **honest** if each σ_i^* is honest. The terminal beliefs $\hat{\beta}^* = (\hat{\beta}_i^*)_{i \in I}$ and their associated terminal hierarchies of beliefs $\hat{\delta}^* = (\hat{\delta}_i)_{i \in I}$ are **honest** if $\hat{\beta}^*$ is consistent with the honest strategy profile σ^*. From now on, each σ^*, $\hat{\beta}^*$, and $\hat{\delta}^*$ with superscript $*$ represents the honest strategy profile, a honest terminal belief profile, and a honest profile of terminal hierarchies of beliefs, respectively.

[9] Belief-dependent outside options are beyond the scope of this paper. Agents endogenously form posterior beliefs by conditioning ex post on (imperfectly) observed behavior in the Bayesian game. However, outside options are not actions chosen within the Bayesian game. Hence, it is not clear how outside options can be associated with endogenous beliefs. Extending the model would require a Bayesian game with actions linked to the outside options. Moreover, this would require a comprehensive description of the exogenous and endogenous restrictions on what the agents observe, both within and outside the mechanism.

[10] The set $(\overline{Y} \times M)^\Theta$ denotes the space of mappings from Θ to $\overline{Y} \times M$.

Remark 4.1. While there is a unique honest strategy profile σ^*, there might be multiple terminal beliefs $\hat{\beta}^*$ consistent with σ^*. To see this, say $[\theta_i, \theta_i', y_i, h_i]$ is unreachable if for each $\theta_{-i} \in \Theta_{-i}$, $(y_i, h_i) \notin \text{Supp}\,(\mathcal{M}(\theta_i', \theta_{-i}))$. Note, if the terminal information set $[\theta_i, \theta_i', y_i, h_i]$ is unreachable, then consistency imposes no restriction on the belief $\hat{\beta}_i^*([\theta_i, \theta_i', y_i, h_i])$.

4.1. Believability

Under the canonical Bayesian game $(\Gamma(\mathcal{M}), \mu)$, each agent i receives a message $h_i \in M_i$. In principle, this message may not coincide with the actual posterior beliefs that i has.[11] Nevertheless, for a certain class of extended direct mechanisms, the messages coincide with the posterior beliefs that the agents have along the honest strategy path. We will be interested in characterizing that class of extended direct mechanisms.

To do so, observe that the prior μ and the honest strategy profile σ^* induce a probability measure $\phi \in \Delta(\Theta \times \overline{Y} \times M)$ defined by $\phi(\theta, y, h) = \mu(\theta) \cdot \mathcal{M}(\theta)(y, h)$. The value $\phi(\theta, y, h)$ is the probability that the type profile is θ and chance selects $(y, h) \in \overline{Y} \times M$ given that all agents report truthfully. Call ϕ the **ex-ante probability measure** induced by $\mathcal{M}$.

To characterize when the suggested hierarchies coincide with the agents' beliefs, it is convenient to consider the canonical homeomorphism between the spaces H_i and $\Delta(\Theta_{-i} \times H_{-i})$. (See Brandenburger and Dekel (1993).) This homeomorphism identifies each hierarchy $h_i \in H_i$ with a unique probability measure $h_i^\infty \in \Delta(\Theta_{-i} \times H_{-i})$ that represents the beliefs that i has about types and hierarchies of beliefs of the other agents. (See Appendix A.) Call h_i^∞ the **extension** of h_i. The finite set of messages $M = \prod_{i \in I} M_i \subseteq H$ is **belief-closed** if, for each $h_i \in M_i$, its associated extension satisfies $h_i^\infty(\Theta_{-i} \times M_{-i}) = 1$.

Definition 4.2. Fix an extended direct mechanism $\mathcal{M}$ with support $\overline{Y} \times M$ and let ϕ be the ex-ante probability measure induced by $\mathcal{M}$. Say $\mathcal{M}$ satisfies **believability (BLV)** if, for each $(\theta_i, \theta_{-i}) \in \Theta$, $y_i \in \overline{Y}_i$, and $(h_i, h_{-i}) \in M$,

$$h_i^\infty(\theta_{-i}, h_{-i}) \cdot \text{marg}_{\Theta_i \times Y_i \times M_i}\, \phi(\theta_i, y_i, h_i) = \text{marg}_{\Theta \times Y_i \times M}\, \phi(\theta_i, \theta_{-i}, y_i, h_i, h_{-i}).$$

The believability condition is an independence property on the ex-ante distribution of posterior beliefs induced by $\mathcal{M}$. It states that the hierarchy-message h_i is sufficient for updating and no further information is gained from (θ_i, y_i). Lemma C.3 shows that if $\mathcal{M}$ satisfies believability, then M is belief-closed.

Say $\hat{\rho}_i \in \hat{\mathcal{P}}_i$ is **reached on the honest equilibrium path**, if there is some $\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}$ such that $\mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid \sigma^*) > 0$. (See Appendix A.2 for the definition of $\mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid \sigma^*)$.) Notice, if a terminal information set $\hat{\rho}_i$ is not reached on the honest equilibrium path, either i did not provide a truthful report or $\hat{\rho}_i$ is unreachable.

Lemma 4.1. *Fix an extended direct mechanism $\mathcal{M}$ satisfying BLV, fix $(\theta_i, y_i, h_i) \in \Theta_i \times \overline{Y}_i \times M_i$, and let $\hat{\delta}^*$ be a profile of honest hierarchy mappings.*

(i) If $[\theta_i, \theta_i, y_i, h_i] \in \hat{\mathcal{P}}_i$ is reached on the honest equilibrium path, then $\hat{\delta}_i^*([\theta_i, \theta_i, y_i, h_i]) = h_i$.

(ii) If types are independent, and $[\theta_i, \theta_i, y_i, h_i] \in \hat{\mathcal{P}}_i$ is reached on the honest equilibrium path, then $\hat{\delta}_i^*([\theta_i', \theta_i, y_i, h_i]) = h_i$ for each type $\theta_i' \in \Theta_i$.

Part (i) states that, under believability, the message that i receives on i's honest path coincides with i's beliefs derived by Bayes rule. Part (ii) adds that, if types are independent, i's message coincides with i's posterior beliefs, even if i's report is not truthful.

4.2. Bayesian incentive compatibility

Bayesian incentive compatibility states that there are honest beliefs $\hat{\beta}^*$ for which each type θ_i prefers to report θ_i over any other report θ_i'.

Definition 4.3. Call an extended direct mechanism $\mathcal{M}$ **Bayesian incentive compatible (BIC)** if there exist honest terminal beliefs $\hat{\beta}^*$ such that, for each $\theta_i, \theta_i' \in \Theta_i$,

$$\mathbb{E}u_i(\theta_i, \theta_i \mid \sigma_{-i}^*, \hat{\beta}_i^*) \geq \mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma_{-i}^*, \hat{\beta}_i^*).$$

BIC implies that the honest profile (σ^*, β^*) constitutes a Bayesian equilibrium of $(\Gamma(\mathcal{M}), \mu)$. Notice that, in principle, (σ^*, β^*) may be a Bayesian equilibrium, even if there are honest terminal beliefs β^{**} so that (σ^*, β^{**}) is not a Bayesian equilibrium. Lemma C.6 shows that this is not the case if $\mathcal{M}$ satisfies believability. Believability implies that β^* and β^{**} agree on reachable terminal information sets. While β^* and β^{**} can differ on unreachable terminal information sets, the difference has no impact on expected payoffs. Consequently, (σ^*, β^{**}) is also a Bayesian equilibrium.

[11] To see this, consider a trivial extended direct mechanism that selects the same outcome and profile of messages irrespectively of the agents' reports. Under that mechanism, the agents' posterior beliefs must align with their prior beliefs, even if the message suggests otherwise.

5. The revelation principle

Fix an arbitrary extensive form Γ and a Bayesian equilibrium $(\sigma, \hat{\beta})$ of (Γ, μ). The revelation principle relies on the construction of an extended direct mechanism $\mathcal{M}$ that "represents" the original Bayesian equilibrium $(\sigma, \hat{\beta})$.

To construct such mechanism, let $\hat{\delta}$ be the terminal hierarchy mappings induced by $\hat{\beta}$ and write $\hat{\mathcal{P}}[\theta, y, h]$ for the set of terminal information sets consistent with type profile θ, collective material outcome y, and hierarchy profile h in the original Bayesian game (Γ, μ). So,

$$\hat{\mathcal{P}}[\theta, y, h] := \left\{\hat{\rho} \in \hat{\mathcal{P}} : \ \text{proj}_{\Theta}(\hat{\rho}) = \theta, \ \text{proj}_{Y}(\hat{\rho}) = y, \ \text{and } \hat{\delta}(\hat{\rho}) = h\right\}.$$

In addition, write $\mathbb{P}(\hat{\rho} \mid \theta, \sigma)$ for the interim probability of reaching $\hat{\rho} = (\hat{\rho}_i)_{i \in I}$ given the type profile θ and strategy profile σ. (See Appendix A for its calculation.)

Definition 5.1. Fix a Bayesian equilibrium $(\sigma, \hat{\beta})$ of (Γ, μ) and let $\hat{\delta}$ be the hierarchy mapping induced by $\hat{\beta}$. Write $\overline{Y} \subseteq Y$ for the range of $\text{proj}_Y : \hat{\mathcal{P}} \to Y$ and $M \subseteq H$ for the range of $\hat{\delta} : \hat{\mathcal{P}} \to H$. The **extended direct mechanism induced by** $(\sigma, \hat{\beta})$ is a mapping $\mathcal{M} : \Theta \to \Delta(Y \times H)$ so that, for each $(\theta, y, h) \in \Theta \times \overline{Y} \times M$,

$$\mathcal{M}(\theta)(y, h) := \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta, y, h]} \mathbb{P}(\hat{\rho} \mid \theta, \sigma).$$

Lemma D.3 shows that, for each $\theta \in \Theta$, $\mathcal{M}(\theta)(\cdot)$ is a well defined probability measure with support in $\overline{Y} \times M$.

It will be useful to differentiate the expected utility functions of the original Bayesian game (Γ, μ) from the canonical Bayesian game $(\Gamma(\mathcal{M}), \mu)$. Toward that end, write $\mathbb{E}u_i(\cdot \mid \sigma_{-i}, \hat{\beta})$ (resp. $\mathbb{E}u_i^*(\cdot \mid \sigma_{-i}^*, \hat{\beta}^*)$) for i's expected utility under $(\sigma_{-i}, \hat{\beta})$ in the original Bayesian game (resp. i's expected utility under $(\sigma_{-i}^*, \hat{\beta}^*)$ in the canonical Bayesian game). Similarly, write $\mathbb{E}\pi(\sigma, \hat{\beta})$ (resp. $\mathbb{E}\pi^*(\sigma^*, \hat{\beta}^*)$) for the designer's expected payoff under $(\sigma, \hat{\beta})$ in the original Bayesian game (resp. the designer's expected payoff under $(\sigma^*, \hat{\beta}^*)$ in the canonical Bayesian game under $(\sigma^*, \hat{\beta}^*)$).

Theorem 5.1. ***(The revelation principle)*** *Let $(\sigma, \hat{\beta})$ be a Bayesian equilibrium of (Γ, μ). The extended direct mechanism $\mathcal{M}$ induced by $(\sigma, \hat{\beta})$ satisfies BLV and BIC. Moreover:*

(i) Each honest profile $(\sigma^*, \hat{\beta}^*)$ is a Bayesian equilibrium of $(\Gamma(\mathcal{M}), \mu)$.

(ii) For each $\theta_i \in \Theta_i$, $\mathbb{E}u_i^*(\theta_i, \sigma_i^* \mid \sigma_{-i}^*, \hat{\beta}^*) = \mathbb{E}u_i(\theta_i, \sigma_i \mid \sigma_{-i}, \hat{\beta})$.

(iii) $\mathbb{E}\pi^*(\sigma^*, \hat{\beta}^*) = \mathbb{E}\pi(\sigma, \hat{\beta})$.

Part (i) states that each honest profile $(\sigma^*, \hat{\beta}^*)$ is a Bayesian equilibrium of the canonical Bayesian game. Part (ii) states that each type θ_i has the same interim expected utility from $(\sigma^*, \hat{\beta}^*)$ as it does in the original Bayesian equilibrium $(\sigma, \hat{\beta})$. Part (iii) states that the designer gets the same ex-ante expected payoffs in both equilibria.

The revelation principle establishes a tool that simplifies mechanism design problems under partial implementation. It implies that there is no loss of generality in optimizing within the set of extended direct mechanisms that satisfy BLV, BIC, and individual rationality (IR) constraints. If the designer can freely choose the Bayesian equilibrium played, there is nothing to be gained in complex and multi-stage mechanisms. To see this, suppose that the extensive form Γ and an associated Bayesian equilibrium $(\sigma, \hat{\beta})$ constitute a solution of the designer's problem. That is, the pair $(\Gamma, (\sigma, \hat{\beta}))$ maximizes the designer's expected payoff subject to equilibrium conditions and IR constraints. Let $\mathcal{M}$ be the extended direct mechanism induced by (Γ, μ) and $(\sigma, \hat{\beta})$, and let $(\sigma^*, \hat{\beta}^*)$ be a honest profile of the canonical Bayesian game. By revelation principle, $(\sigma^*, \hat{\beta}^*)$ satisfies (i), (ii), and (iii). Notice that (i) implies that $(\sigma^*, \hat{\beta}^*)$ is a Bayesian equilibrium, (ii) implies that $(\sigma^*, \hat{\beta}^*)$ satisfies IR, and (iii) implies that $(\sigma^*, \hat{\beta}^*)$ achieves the designer's optimal expected payoff. Therefore, $(\Gamma(\mathcal{M}), (\sigma^*, \hat{\beta}^*))$ is also a solution of the designer's problem.

5.1. Outline of proof

The proof of Theorem 5.1 can be found in Appendix D. This section presents an outline.

Fix an extensive form Γ and let $(\hat{\mathcal{P}}_i)_{i \in I}$ be the information partitions of (Γ, μ). Consider a Bayesian equilibrium $(\sigma, \hat{\beta})$ of (Γ, μ) and write $\mathbb{P}(\cdot \mid \sigma)$ for the ex-ante distribution on $\hat{\mathcal{P}}$ given σ. Let $\mathcal{M}$ be the extended direct mechanism induced by $(\sigma, \hat{\beta})$. The support of $\mathcal{M}$ is given by $\overline{Y} \times M$ and the ex-ante probability measure of $\mathcal{M}$ is ϕ.

The proof establishes two key identities, derived from the construction of $\mathcal{M}$:

(a) For each $(\theta, y, h) \in \Theta \times \overline{Y} \times M$, $\phi(\theta, y, h) = \mathbb{P}(\hat{\mathcal{P}}[\theta, y, h] \mid \sigma)$.

(b) For each $\theta_i, \theta_i' \in \Theta_i$, $\mathbb{E}u_i^*(\theta_i, \theta_i' \mid \hat{\sigma}_{-i}^*, \hat{\beta}^*) = \sum_{s_i \in S_i} \mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta}) \cdot \sigma_i(\theta_i')(s_i)$.

Identity (a) links ϕ with the ex-ante distribution of types, material outcomes, and hierarchies induced by the equilibrium $(\sigma, \hat{\beta})$ in (Γ, μ). Identity (b) connects the agents' interim expected utilities between the two Bayesian games: θ_i's expected utility from playing

$\sigma_i(\theta_i')$ given the equilibrium $(\sigma, \hat{\beta})$ in (Γ, μ) equals θ_i's expected utility from reporting θ_i' given an honest equilibrium of $(\Gamma(\mathcal{M}), \mu)$. Importantly, identity (b) holds for any choice of honest terminal beliefs $\hat{\beta}^*$ and any report θ_i', including reports that are not truthful.

Parts (ii) and (iii) of the theorem immediately follow from identities (b) and (a), respectively. To show Part (i), it suffices to show that $\mathcal{M}$ satisfies BLV and BIC.

Establishing BLV requires showing

$$h_i^{\infty}(\theta_{-i}, h_{-i}) \cdot \operatorname{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i) = \operatorname{marg}_{\Theta \times \overline{Y}_i \times M} \phi(\theta_i, \theta_{-i}, y_i, h_i, h_{-i}).$$

The key is to use condition (a) to rewrite BLV as a relation between the terminal beliefs $\hat{\beta}$ and the strategy profile σ. Doing so, requires linking the value $h_i^{\infty}(\theta_{-i}, h_{-i})$ with i's terminal beliefs about the set $\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}] := \left\{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i} : \operatorname{proj}_{\Theta_{-i}}(\hat{\rho}_{-i}) = \theta_{-i} \text{ and } \hat{\delta}_{-i}(\hat{\rho}_{-i}) = h_{-i}\right\}$. With this in mind, observe that if $\hat{\rho}_i \in \hat{\mathcal{P}}_i[\theta_i, y_i, h_i]$, then i has hierarchy h_i at $\hat{\rho}_i$, and, consequently, $\hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]) = h_i^{\infty}(\theta_{-i}, h_{-i})$. Moreover, identity (a) implies that

$$\operatorname{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i) = \mathbb{P}(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i} \mid \sigma)$$

and

$$\operatorname{marg}_{\Theta \times \overline{Y}_i \times M} \phi(\theta_i, \theta_{-i}, y_i, h_i, h_{-i}) = \mathbb{P}(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}] \mid \sigma).$$

Therefore, BLV holds if and only if

$$\hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}])) \cdot \mathbb{P}(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i} \mid \sigma) = \mathbb{P}(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}] \mid \sigma)$$

holds for each $\hat{\rho}_i \in \hat{\mathcal{P}}_i[\theta_i, y_i, h_i]$. This equation indicates a relation between σ and $\hat{\beta}$. The proof shows that this condition follows from consistency of $(\sigma, \hat{\beta})$, and, as a result, BLV is satisfied.

Finally, observe that BIC follows immediately from identity (b). As in the textbook revelation principle, if θ_i does not gain by playing $\sigma(\theta_i')$ in (Γ, μ), then θ_i does not gain by reporting θ_i' in $(\Gamma(\mathcal{M}), \mu)$. The presence of belief-dependent preferences does not change this fundamental idea.

6. Auction design under belief-dependent preferences

This section applies the revelation principle to study optimal auction design in a private values setting with the belief-dependent preferences of Section 2. The analysis provides a characterization of the revenue achievable and sufficient conditions for revenue maximization. In addition, it describes the optimal way to allocate the object and reveal information when agents have different types of image concerns.

Fix an extended direct mechanism $\mathcal{M}$ with support $\overline{Y} \times M$. As in standard models, we need to describe how different reports change the probability of wining the object and the payment made to the auctioneer. With this in mind, write $W_i := \left\{(x_j, t_j)_{j \in I} \in Y : x_i = 1\right\}$ for the set of collective material outcomes where bidder i wins the object and write

$$Q_i(\theta_i) := \sum_{\theta_{-i} \in \Theta_{-i}} \operatorname{marg}_{\Theta_{-i}} \mu(\theta_{-i}) \cdot \operatorname{marg}_{Y} \mathcal{M}(\theta_i, \theta_{-i})(W_i),$$

for the probability of bidder i winning the object after a report θ_i assuming that the other agents truthfully report their types.[12] Write

$$T_i(\theta_i) := \sum_{(x_i, t_i) \in \overline{Y}_i} \sum_{\theta_{-i} \in \Theta_{-i}} t_i \cdot \operatorname{marg}_{\Theta_{-i}} \mu(\theta_{-i}) \cdot \operatorname{marg}_{Y_i} \mathcal{M}(\theta_i, \theta_{-i})(x_i, t_i),$$

for the expected transfer for i after a report θ_i, assuming that the other agents truthfully report their types. So, if there is truthful reporting in $\mathcal{M}$, the auctioneer's **expected revenue** from i is $\operatorname{Rev}_i(\mathcal{M}) := \sum_{\theta_i \in \Theta_i} T_i(\theta_i) \cdot \overline{\mu}_i(\theta_i)$.

Because the setting has belief-dependent preferences, characterizing the agents' incentives, requires describing how different reports influence the psychological sub-utility. Towards that end, write

$$F_i(\theta_i) := \sum_{h_i \in M_i} \sum_{\theta_{-i} \in \Theta_{-i}} f_i(h_i) \cdot \operatorname{marg}_{\Theta_{-i}} \mu(\theta_{-i}) \cdot \operatorname{marg}_{H_i} \mathcal{M}(\theta_i, \theta_{-i})(h_i),$$

for i's expected psychological sub-utility assuming i reports θ_i, other agents truthfully report their types, and i's beliefs about hierarchies coincide with the mechanism's messages.

Call $(Q_i(\cdot), T_i(\cdot), F_i(\cdot))_{i \in I}$ the **outcome mappings** of $\mathcal{M}$. Lemma E.1 shows that, if $\mathcal{M}$ satisfies believability, then the expected utility of a type θ_i that reports θ_i' when others report truthfully is $U_i(\theta_i, \theta_i') := Q_i(\theta_i') \cdot \theta_i - T_i(\theta_i') + F_i(\theta_i')$. Recall that agents have a value option of zero value. So, $\mathcal{M}$ is **individually rational** (IR) if $U_i(\theta_i, \theta_i) \geq 0$ for each $\theta_i \in \Theta_i$, i.e., if each type has an incentive to participate in the mechanism under the honest profile.

[12] Notice, since types are independent, this probability depends on i's report but not on i's realized type. Similarly, the expected transfers and messages depend on i' report but do not depend on i's realized type.

6.1. Revenue characterization

Bergemann and Pesendorfer (2007) (B&P) study the finite type-space version of auction design in settings with private valuations and standard preferences. Hence, they serve as a benchmark to compare how the auctioneer's expected revenue changes in the presence of belief-dependent preferences.

It will be useful to review that benchmark. Towards that end, fix an agent $i \in I$ and write $\Theta_i = \{\theta_i^1, \dots, \theta_i^K\}$ with $\theta_i^{k+1} > \theta_i^k$. B&P show that the virtual valuations capture the maximum revenue that can be extracted under BIC and IR constraints. They define the **virtual valuations** for agent i as

$$v_i(\theta_i^k) := \begin{cases} \theta_i^k - (\theta_i^{k+1} - \theta_i^k) \cdot \frac{1 - \sum_{\ell=1}^{k} \overline{\mu}_i\left(\theta_i^\ell\right)}{\overline{\mu}_i(\theta_i^k)} & \text{if } k < K \\ \theta_i^K & \text{if } k = K. \end{cases}$$

These values are the discrete version of the virtual valuations for continuous-type environments defined by Myerson (1981).

Example 6.1. Let $\Theta_i = \{1, \dots, K\}$ and let each value is equally likely according to $\overline{\mu}_i \in \Delta(\Theta_i)$. Then, $v_i(\theta_i^k) = 2k - K$. Hence, for this type distribution, i virtual values are increasing, low types have negative virtual values, and high types have positive virtual values.

In the context of belief-dependent preferences, the same virtual values constitute one part of the revenue that can be extracted by the auctioneer.

Proposition 6.1. *Fix an extended direct mechanism $\mathcal{M}$ that satisfies BLV and BIC; let $(Q_i(\cdot), T_i(\cdot), F_i(\cdot))_{i \in I}$ be its associated outcome mappings.*

(i) For each $i \in I$, $Q_i(\cdot)$ is weakly increasing.

(ii) There exists $(\tilde{v}_i^1, \dots, \tilde{v}_i^K) \in \mathbb{R}^k$ such that:

- *if $Q_i(\theta_i^k) = 0$ or $k = K$, then $\tilde{v}_i^k = v_i(\theta_i^k)$;*
- *if $Q_i(\theta_i^k) \neq 0$ and $k < K$, then $v_i(\theta_i^k) \geq \tilde{v}_i^k \geq \theta_i^k - (\theta_i^{k+1} - \theta_i^k) \cdot \frac{1 - \sum_{\ell=1}^{k} \overline{\mu}_i(\theta_i^\ell)}{\overline{\mu}_i(\theta_i^k)} \cdot \frac{Q_i(\theta_i^{k+1})}{Q_i(\theta_i^k)}$.*

Moreover,

$$\mathrm{Rev}_i(\mathcal{M}) = \sum_{k=1}^{K} Q_i(\theta_i^k) \cdot \tilde{v}_i^k \cdot \overline{\mu}_i(\theta_i^k) - U_i(\theta_i^1, \theta_i^1) + \sum_{k=1}^{K} F_i(\theta_i^k) \cdot \overline{\mu}_i(\theta_i^k).$$

Proposition 6.1 extends Lemma 1 in B&P to environments with belief-dependent preferences. The result has two parts. The first part mirrors the monotonicity condition from standard settings. Despite the presence of belief-dependent preferences, an allocation rule is consistent with BIC only if types with higher valuation are weakly more likely to win the object. The second part characterizes expected revenue for extended direct mechanisms that satisfies BLV and BIC. The revenue extracted from i has three components: The first term is the ex-ante expectation of $Q_i(\theta_i^k) \cdot \tilde{v}_i^k$—which has a close relation with i's ex-ante virtual welfare.[13] The second term is the utility of i's lowest type. The third term is i's ex-ante expected psychological sub-utility. While the first two components appear in classical results, the third component is new and emerges only in contexts with belief-dependent preferences. The pivotal argument in the proof is that the additive separability of the psychological sub-utility enables the problem to be restructured into a classical framework for some auxiliary transfers. Moreover, Lemma E.3 uses this transformation to provide a converse to Proposition 6.1.

The key insight of Proposition 6.1 is that the auctioneer completely captures the agents' ex-ante psychological sub-utility. So, if $f_i(\cdot)$ takes non-negative values, agent i's ex-ante belief-dependent gains are fully extracted by the designer.

The full extraction of the agents' psychological sub-utility hinges on the fact that, in equilibrium, the designer infers the value of i's posterior hierarchy and is able to fully charge for it. This is possible due two key assumptions: (1) the agents' utility function is additively separable in i's material and psychological sub-utility, and (2) the agents' psychological sub-utility $f_i(h_i)$ does not directly depend on θ_i. So, unlike the agent's unknown value of the object θ_i, the auctioneer exactly knows i's valuation for each hierarchy h_i. Consequently, agents obtain the same information rents from their material sub-utility but are not able to profit from their psychological sub-utility.

Notice, if f_i takes negative values, then the auctioneer needs to cover the agents' belief-dependent costs to encourage the agents' participation in the mechanism.

6.2. Revenue maximization

An extended direct mechanism is **implementable** if it satisfies BLV, BIC, and IR. Write IMP for the set of implementable extended direct mechanisms. For each $\mathcal{M} \in \mathrm{IMP}$ write $\mathrm{Rev}(\mathcal{M}) := \sum_{i \in I} \mathrm{Rev}_i(\mathcal{M})$. By the revelation principle, a mechanism $\mathcal{M} \in \mathrm{IMP}$ is

[13] The indeterminacy of the numbers $(\tilde{v}_i^1, \dots, \tilde{v}_i^K) \leq (v_i(\theta_i^1), \dots, v_i(\theta_i^K))$ arises due to the discrete nature of the type space. In the same way as in B&P, the BIC constraints characterize incentive-compatible transfers within an interval. However, this indeterminacy is not relevant for revenue maximization, as $(\tilde{v}_i^1, \dots, \tilde{v}_i^K) = (v_i(\theta_i^1), \dots, v_i(\theta_i^K))$ if transfers are carefully tailored. (See Lemma E.2.)

revenue-maximizing if $\text{Rev}(\mathcal{M}) \geq \text{Rev}(\mathcal{M}')$ for each $\mathcal{M}' \in \text{IMP}$. This section identifies features of revenue-maximizing mechanisms. For each $\mathcal{M} \in \text{IMP}$, write

$$\text{VW}(\mathcal{M}) := \sum_{i \in I} \sum_{\theta_i \in \Theta_i} Q_i(\theta_i) \cdot v_i(\theta_i) \cdot \overline{\mu}_i(\theta_i),$$

for the mechanism's **virtual welfare**. The virtual welfare is the sum of the agents' ex-ante expected virtual values. Additionally, write

$$\text{PW}(\mathcal{M}) := \sum_{i \in I} \sum_{\theta_i \in \Theta_i} F_i(\theta_i) \cdot \overline{\mu}_i(\theta_i),$$

for the mechanism's **psychological welfare**. The psychological welfare is the sum of the agents' ex-ante expected psychological sub-utility. A mechanism $\mathcal{M}$ has **maximally-compatible transfers** if

$$T_i(\theta_i^k) = \begin{cases} \theta_i^1 \cdot Q_i(\theta_i^1) + F_i(\theta_i^1) & \text{if } k = 1 \\ \theta_i^k \cdot Q_i(\theta_i^k) - \sum_{\ell=1}^{k-1} (\theta_i^{\ell+1} - \theta_i^{\ell}) \cdot Q_i(\theta_i^{\ell}) + F_i(\theta_i^k) & \text{if } k > 1. \end{cases} \tag{1}$$

The virtual welfare, the psychological welfare, and mechanisms with maximally-compatible transfers play an important role in identifying revenue-maximizing mechanisms. Lemma E.2 applies the revenue characterization to show that $\text{Rev}(\mathcal{M}) \leq \text{VW}(\mathcal{M}) + \text{PW}(\mathcal{M})$ for each $\mathcal{M} \in \text{IMP}$. Moreover, it shows that the inequality binds if $\mathcal{M}$ has maximally-compatible transfers. Consequently, we obtain the following criteria for revenue maximization:

Corollary 6.1. *Fix a mechanism $\mathcal{M} \in \text{IMP}$. If $\mathcal{M}$ has maximally-compatible transfers and*

$$\text{VW}(\mathcal{M}) + \text{PW}(\mathcal{M}) = \sup_{\mathcal{M}' \in \text{IMP}} \left(\text{VW}(\mathcal{M}') + \text{PW}(\mathcal{M}') \right),$$

then $\mathcal{M}$ is revenue-maximizing.

In standard auction models, revenue maximization is achieved by maximizing the allocation's virtual welfare while properly adjusting the agents' transfers. That is, when each $F_i(\cdot) = 0$, the mapping $Q_i(\cdot)$ determines maximum compatible transfers that maximize revenue. (See Myerson (1981) and Bergemann and Pesendorfer (2007).)

Corollary 6.1 extends this result to belief-dependent preferences: Revenue maximization is achieved by finding mappings $Q_i(\cdot)$ and $F_i(\cdot)$ that maximize the sum of the virtual welfare and psychological welfare, and then adjusting the transfers according to Equation (1).

6.3. Optimal auctions

This section uses Corollary 6.1 to identify the revenue-maximizing auctions in environments with expectation-based, sophisticated-type, unsophisticated-type image concerns, and the simple image concerns in Example 2.6.

First we focus on expectation-based, sophisticated-type, and unsophisticated-type image concerns. We analyze the setting where $Y_i = \{(x_i, t_i) \in \{0,1\} \times \mathbb{R}\}$. So, there are no individual restrictions in allocations and transfers. In addition, two standard assumptions are imposed. First, it is assumed that the agents are ex-ante symmetric, meaning that $\Theta_i = \Theta_j$ and $\overline{\mu}_i = \overline{\mu}_j$ for all $i, j \in I$. Second, it is assumed that the associated virtual valuations $v_i : \Theta_i \to \mathbb{R}$ are strictly increasing.[14]

Definition 6.1. An extended direct mechanism $\mathcal{M}$ has a **virtual-value cutoff** if it satisfies the following:

(i) If $\theta \in \Theta$ is such that, for some i $v_i(\theta_i) > 0$, then $\text{marg}_Y \mathcal{M}(\theta)(\cup_{i \in I} W_i) = 1$.

(ii) If $\theta \in \Theta$ is such that, for some i, either $v_i(\theta_i) < 0$ or $\theta_i < \max_{j \in I \setminus \{i\}} \theta_j$, then $\text{marg}_Y \mathcal{M}(\theta)\left(W_i\right) = 0$.

Mechanisms with a virtual value cutoff allocate the object whenever some agent has a strictly positive virtual value (Condition (i)). Moreover, the mechanism allocates the object only to agents with non-negative virtual values that have the highest valuations (Condition (ii)). So, if all agents have negative virtual valuations, the object is kept by the auctioneer. Lemma E.5 shows that, under ex-ante symmetry and increasing virtual valuations, $\mathcal{M} \in \text{IMP}$ maximizes virtual welfare if and only if it has a virtual-value cutoff.

Proposition 6.2. *Assume that the agents are ex-ante symmetric and that virtual values are strictly increasing. The following hold for settings with standard preferences, expectation-based, sophisticated-type, and unsophisticated-type image concerns:*

(i) There exist an implementable revenue-maximizing mechanism.

[14] These are standard assumptions in the literature. Absent the assumption of ex-ante symmetry, the object may not be allocated to the agents with highest valuations. Absent of the assumption of increasing virtual values, a partial characterization of revenue-maximizing mechanisms can be obtained by using ironing techniques. (See Bergemann and Pesendorfer (2007).)

(ii) Any implementable revenue-maximizing mechanism has a virtual-value cutoff.

Proposition 6.2 shows that, as with standard preferences, in environments with expectation-based, sophisticated-type, and unsophisticated-type image concerns, revenue-maximizing mechanisms have a virtual-value cutoff. Hence, the use of reserve prices, which preclude agents with negative virtual valuations from winning the object, continues to be optimal in settings involving these types of image concerns.

While the optimal allocation does not depend on whether the agents have expectation-based, sophisticated-type, or unsophisticated-type image concerns, the optimal way to reveal information is sensitive to the type of image concerns.

Proposition 6.3. *Assume that the agents are ex-ante symmetric and that virtual values are strictly increasing.*

(i) If agents have standard preferences or expectation-based image concerns, the information revealed to the agents does not change the auctioneer's expected revenue. In particular, there are both revenue-maximizing mechanisms that publicly reveal the agents' types and revenue-maximizing mechanisms that conceal the agents' types.

(ii) If agents have sophisticated-type image concerns, the expected revenue is sensitive to the information revealed by the mechanism: Any revenue-maximizing mechanism publicly reveals whether the agents' valuations achieve the threshold b.

(iii) If agents have unsophisticated-type image concerns, the expected revenue is sensitive to the information revealed by the mechanism: Any revenue-maximizing mechanism conceals whether the agents' valuations achieve the threshold b.

Proposition 6.2 shows that maximizing revenue requires revealing different information based on the specific type of image concern.[15]

Under expectation-based image concerns, the agents' psychological sub-utility is linear in the agents' beliefs. As a result, the agents' psychological welfare is invariant under posterior spreads. Thus, from an ex-ante perspective, the agents' are indifferent between revealing or concealing their valuations. Hence, to maximize revenue, the auctioneer can use an auction with a reserve price and a participation fee. For instance, the mechanism may use the rules of an English auction, a Dutch auction, or a sealed-bid auction. Despite the fact that these different auctions reveal different information about the agents' bids, they do not affect the auctioneer's expected revenue. The key is to select a reserve price in a way that types with negative virtual valuations never win the object. Furthermore, given that agents forfeit their psychological sub-utility if they do not participate, the auctioneer can use the participation fee to extract the psychological sub-utility of types with negative virtual valuations.[16]

Under sophisticated-type image concerns, the agents are better off when they can publicly reveal their valuations. Concealing the agents' valuations would only preclude the types above the threshold b from receiving their psychological reward. Hence, the optimal auction should publicly reveal whether the agents' valuations achieve the threshold. To do so, the auctioneer can use an auction with public bids, a reserve price, and the option for bidders to submit a donation to the auctioneer. For example, the auctioneer can use the rules of a first-price auction, provided that bids are publicly revealed at the end of the auction. The reserve price is chosen so that types with negative virtual valuations bid below the reserve price, and thus, never receive the object. Types that are above the threshold b and have positive virtual values use the public bids to credibly show that they are above the threshold. Types above the threshold b with negative virtual values cannot use public bids to signal their values, as lower types would mimic them. Instead, they provide a public donation—equal to the psychological reward a—to credibly show that they are above the threshold. Notice, while all types with negative virtual values are indifferent between providing the public donation or not, only those above the threshold do so.

Under unsophisticated-type image concerns, the agents are better off when they can conceal their valuations. Revealing the valuations would only preclude the types below the threshold b from receiving their psychological reward. Hence, the optimal auction should conceal whether the agents achieve the threshold. To do so, the auctioneer can use an auction with sealed bids, a reserve price, and a participation fee. For instance, the auctioneer can use the rules of the first-price auction provided that bids are never revealed. The reserve price is chosen so that types with negative virtual valuations bid below the reserve price, and thus, never receive the object. Since the bids are not revealed, each agent i only observes whether i wins the object. In particular, there is never common belief that any agent is below threshold b, so all the participants receive their psychological reward a. Since agents forfeit their psychological reward if they do not participate, the auctioneer can extract the agents' psychological sub-utility by charging a participation fee equal to a.

Notice, under these three types of image concerns, revenue maximization can always be achieved by direct mechanisms that either fully reveal or fully conceal the agents' reports (provided that agents report truthfully). However, this does not hold for other types of image concerns—as revenue maximization can require mechanisms that partially reveal information about the agents' types. For instance, consider the simple image concerns described in Example 2.6. If agents report truthfully, direct mechanisms that either fully reveal or fully conceal the agents' reports do not maximize revenue.

Lemma 6.1. *Consider the auction setting described in Example 2.6.*

(i) There exists a mechanism and an individually rational Bayesian equilibrium thereof, that achieves an expected revenue of 6.

[15] Appendix E.1 provides a formal definition of the terms "publicly revealing" and "concealing."

[16] This result extends that in Bos and Pollrich (2022). They showed that, under expectation-based image concerns, reserve prices and participation fees can be used to maximize revenue among the set of symmetric mechanisms and symmetric equilibria. The result here does not impose the symmetry restrictions.

(ii) The expected revenue achieved by any individually rational Bayesian equilibrium of any mechanism is at most 6.
(iii) If the honest strategy profile is an individually rational Bayesian equilibrium of a direct mechanism that either fully reveals or fully conceals Ann's type, then the auctioneer's expected revenue is at most 5.

Lemma 6.1 follows from observing that any mechanism that maximizes revenue must maximize the ex-ante psychological welfare of Ann. Achieving this goal requires a mechanism that partially reveals Ann's type to Bob. In particular, it requires a mechanism that only sends hierarchy messages h_A that satisfy $\mathbb{E}_A[\mathbb{E}_B[\theta_A] \mid h_A] \in \{0,4\}$.

Fig. 6.1 provides geometric intuition. Notice, by the law of iterated expectations, any mechanism satisfying BLV induces a distribution of values of $\mathbb{E}_A[\mathbb{E}_B[\theta_A] \mid h_A]$ with a mean of 3 (the ex-ante expectation of θ_A) and support in $[\min \Theta_A, \max \Theta_A] = [0,6]$. Notice that $f_A(h_A) = 4$ if $\mathbb{E}_A[\mathbb{E}_B[\theta_A] \mid h_A] \geq 4$ and equals 0 otherwise. Thus, by standard concavification methods, the maximum ex-ante psychological sub-utility is given by the concave envelope (the dashed blue line) of f_A (the solid red line) evaluated at the ex-ante expectation 3. Moreover, the optimal distribution of Ann's second order expectation must have support $\{0,4\}$.

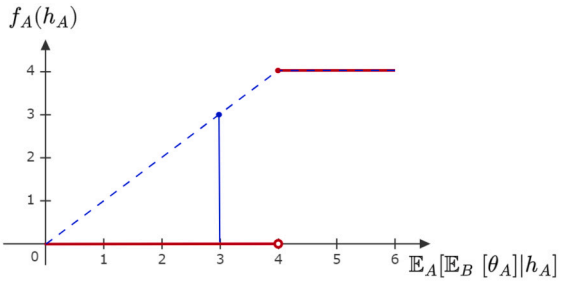

Fig. 6.1. Concave envelope of f_A. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

7. Discussion

7.1. Belief-dependent utilities as a reduced form

The communication revelation principle in Myerson (1982) can be indirectly applied to solve problems with belief-dependent preferences, if the utility function can be written as the expected utility of some Bayesian equilibrium of some after-game. That is, the after-game can be used as an instrument to simulate the preferences about the posterior beliefs.

However, there are some belief-dependent preferences where this is not possible. Examples include unsophisticated-type image concerns (see Lemma F.2), temptation and self-control (Gul and Pesendorfer, 2001), ego utility (Köszegi, 2006), and shame (Dillenberger and Sadowski, 2012).[17] The impossibility arises because any belief-dependent preferences that capture i's expected utility of some after-game must be weakly convex in i's hierarchies of beliefs. The key is that, if such belief-dependent preferences are induced by a standard game, more information cannot make an agent i worse off unless other players know i has more information.

To show the result, we focus on after-games that can depend on the material outcomes of the mechanisms; so at the end of the mechanism, the collective material outcome is publicly observed.[18] Fix an environment (I, Θ, Y). An **after-game** $\mathcal{G}$ is a tuple $(A_i, v_i)_{i \in I}$ where A_i is the measurable space of i's pure strategies and the measurable mapping $v_i : \Theta \times Y \times \prod_{j \in I} A_j \to \mathbb{R}$ is i's utility function. Note that $\mathcal{G}$ is *standard* in the sense that the utility functions only depend on actions, material outcomes, and types, but not beliefs.

Consider the Bayesian after-game induced by $\mathcal{G}$. A **strategy** of agent i in this Bayesian game is a measurable mapping $\alpha_i : \Theta_i \times Y \times H_i \to \Delta(A_i)$. Call $(\alpha_i)_{i \in I}$ a **Bayesian equilibrium** of $\mathcal{G}$, if for each $(\theta_i, y, a_i') \in \Theta_i \times Y \times A_i$,

$$\int_{A_i} \hat{v}_i(\theta_i, y, a_i \mid h_i, \alpha_{-i}) \, d\alpha_i(\theta_i, y, h_i) \geq \hat{v}_i(\theta_i, y, a_i' \mid h_i, \alpha_{-i}),$$

where

$$\hat{v}_i(\theta_i, y, a_i \mid h_i, \alpha_{-i}) := \int_{\Theta_{-i} \times H_{-i}} \int_{A_{-i}} v_i(\theta_i, \theta_{-i}, y, a_i, a_{-i}) \, d\alpha_{-i}(\theta_{-i}, y, h_{-i}) \, dh_i^{\infty}.$$

[17] This list draws from Lipnowski and Mathevet (2018). The impossibility result also holds when preferences depend on both posterior and prior beliefs. Hence, further examples include stubbornness (Lipnowski and Mathevet, 2018) and the two-period version of loss aversion (Köszegi and Rabin, 2009).

[18] This assumption is without loss. The result extents to environments where the collective material outcome is not fully observed. Such analysis requires extending i's first-order domain of uncertainty to $X_i^1 := \Theta_{-i} \times Y_{-i}$, so that i is able to compute beliefs about the actions that other players take. Section 7.5 provides such extension.

That is, α_i optimizes given his type θ_i, his beliefs h_i, and the strategies α_{-i}. The mapping $u_i : \Theta_i \times Y \times H_i \to \mathbb{R}$ is a **reduced-form utility function** of the after-game $\mathcal{G}$ if there exists some Bayesian equilibrium $(\alpha_i)_{i\in I}$ of $\mathcal{G}$ such that

$$u_i(\theta_i, y, h_i) = \int_{A_i} \hat{v}_i(\theta_i, y, a_i \mid h_i, \alpha_{-i})\, d\alpha_i(\theta_i, y, h_i),$$

for each $(\theta_i, y, h_i) \in \Theta_i \times Y \times H_i$.

Proposition 7.1. *If u_i is a reduced-form utility function of some after-game $\mathcal{G}$, then u_i is weakly convex in H_i.*

Proposition 7.1 says that any posterior spread of h_i makes i weakly better off in a reduced-form utility function. So, in environments with after-games, more information cannot make an agent i worse off unless other players know i has more information.

7.2. Material outcomes

The set Y represents the collective material outcomes. These outcomes are exogenous to the problem and do not include the agents' hierarchies of beliefs. This paper assumes that $Y \subseteq \prod_{i\in I} Y_i$, where Y_i represents the dimensions of Y that, for exogenous reasons, i must observe. For instance, in an auction, each agent must observe whether or not he wins the object and his payment. However, the designer may have the capacity to conceal the material outcomes of the other agents. (The description of what the agents observe is irrelevant for in standard mechanism design settings, but relevant with belief-dependent preferences.)

The fact that Y is allowed to be a strict subset of $\prod_{i\in I} Y_i$ allows for a rich set of applications. Here are two examples: First, in an auction setting, it is not feasible for all agents to obtain a single object. Thus, the set of material outcomes Y does not contain all elements of $\{\text{win}, \text{loose}\}^{|I|}$. Second, if all agents observe all dimensions of material outcomes, take $Y_i = Y_j$ and let Y be the diagonal set $\{(y_i)_{i\in I} \in \prod_{i\in I} Y_i : y_i = y_j \text{ for each } i, j \in I\}$.

This paper makes the implicit assumption that all dimensions of the material outcomes are observed by at least one agent. However, the model can be easily extended to a case where there are some dimensions of the material outcomes, Y_0, that are not observable by any agent. In this case, the set of material outcomes is a set $Y \subseteq Y_0 \times \prod_{i\in I} Y_i$. All proofs and results hold in this setting as well.

7.3. Finite environments

This paper assumes a finite environment: The set of type profiles Θ and the set of terminal nodes of the extensive-form are both finite. I anticipate that analogous results hold in non-finite environments.

The restriction to a finite environment has two advantages: first, it circumvents the computation of conditional probabilities in zero-probability events; second, it determines the agents' conditional beliefs. By contrast, when there is a continuum of nodes, there are multiple versions of regular conditional probabilities that are consistent with the equilibrium strategies. Thus, extending the model beyond the finite case would require that the designer can choose the agents' posterior beliefs at each terminal node—even though the posterior beliefs directly impact the agents' utility.

7.4. Dynamic mechanisms and refinements of Bayesian equilibrium

Under Bayesian equilibrium, players only choose strategies at the start of the game and so, only maximize their *ex-ante* expected payoff. In dynamic settings, this feature is undesirable as it allows for non-credible threats. The revelation principle here applies to any dynamic refinement of Bayesian equilibrium. Specifically, it applies to refinements that (1) coincide with Bayesian equilibrium in simultaneous-move games, but (2) refine Bayesian equilibria in sequential games. To see this, consider a dynamic refinement $\mathscr{E}$ of Bayesian equilibrium. Suppose $(\sigma, \widetilde{\beta}, \hat{\beta})$ satisfies $\mathscr{E}$, where $\widetilde{\beta}$ is the profile of non-terminal beliefs. So, $(\sigma, \hat{\beta})$ is a particular Bayesian equilibrium—one that satisfies additional requirements. Notice that the revelation principle applies to $(\sigma, \hat{\beta})$. Moreover, each extended direct mechanism $\mathcal{M}$ induces a simultaneous-move game $\Gamma(\mathcal{M})$. So, the honest profile $(\sigma^*, \hat{\beta}^*)$ associated with $\Gamma(\mathcal{M})$ also satisfies the dynamic refinement $\mathscr{E}$.

The revelation principle implies that dynamic mechanisms do not dominate static ones. This may be surprising given results in the literatures on psychological games and behavioral mechanism design. In particular, some psychological motivations modeled by belief-dependent preferences induce dynamic inconsistencies. Examples include frustration and anger (Battigalli, Dufwenberg, and Smith, 2019b), loss aversion (Köszegi and Rabin, 2007, 2009), and disappointment (Battigalli and Dufwenberg, 2009). (See discussion in Battigalli, Corrao, and Dufwenberg (2019a); Battigalli and Dufwenberg (2022).) At the same time, in environments with dynamically inconsistent agents, dynamic mechanisms can dominate static ones. (See Gershkov, Moldovanu, Strack, and Zhang (2023); Gan (2022)). Thus, one might have expected the same here.

The key is that, here, agents are expected utility maximizers who only care about material outcomes and posterior hierarchical beliefs. This rules out belief-dependent preferences that induce dynamic inconsistencies, e.g. own-plan dependence of "experience utility" and "distortions" in the determination of "decision utility" (Battigalli, Corrao, and Dufwenberg, 2019a). Note that if agents play according to a profile $(\sigma, \hat{\beta})$, the utility of each terminal node is determined by $\hat{\beta}$. Moreover, this utility does not change over the course of the game, precluding any form of dynamic inconsistency. (See Green (1987).)

Because agents here are dynamically consistent, dynamic refinements of Bayesian equilibrium are appropriate solution concepts to describe the agents' behavior. However, that same choice may not be appropriate for other preferences.

Mechanism design in the context of dynamic inconsistencies presents an important avenue for further research. Whether there exists a manageable class of canonical mechanisms for dynamically inconsistent agents (in standard or belief-dependent settings) remains as an open question.

7.5. Beliefs about intentions and material outcomes

The model captures environments in which each agent i has preferences over hierarchical posterior beliefs, where beliefs are about types rather than intentions or material outcomes. Beliefs about intentions—that is, beliefs about actions or strategies that can capture reciprocity, anger, and frustration—are beyond the scope of this paper. However, the analysis can be adapted to incorporate beliefs about material outcomes. Such beliefs are relevant in environments where agents face an after-game where not all agents fully observe the collective material outcome y and the payoffs depend on the type profile θ and the collective material outcome y. For instance, telecommunication companies with private costs might first participate in an auction for electromagnetic spectrum slots and subsequently compete for customers in an after-game. In such an environment, θ represents the firms' private costs, while y represents the payments and allocated slots.

Since the collective material outcome is relevant for the after-game, i's relevant first-order domain of uncertainty is $X_i^1 := \Theta_{-i} \times Y_{-i}$. The framework and the solution concept of Bayesian equilibrium are defined in an analogous way and the believability condition takes the following form:

$$h_i^{\infty}(\theta_{-i}, y_{-i}, h_{-i}) \cdot \operatorname{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i) = \phi(\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i}).$$

An analogous analysis results in a revelation principle for this context.

7.6. Departing from a common prior

This model assumes that the designer's and the agents' beliefs are derived from a common prior $\mu \in \Delta(\Theta)$. However, the approach taken in this paper holds when the agents have different subjective priors $(\mu_i)_{i \in I}$, provided the priors have full support. The priors $(\mu_i)_{i \in I}$ are transparent to the agents, leading the agents to *openly disagree*.

Each extended direct mechanism $\mathcal{M}$ and each subjective prior $\mu_i \in \Delta(\Theta)$ induces i's ex-ante distribution $\phi_i \in \Delta(\Theta \times \overline{Y} \times M)$. In this environment, the believability condition requires

$$h_i^{\infty}(\theta_{-i}, h_{-i}) \cdot \operatorname{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi_i(\theta_i, y_i, h_i) = \operatorname{marg}_{\Theta \times \overline{Y}_i \times M} \phi_i(\theta_i, \theta_{-i}, y_i, h_i, h_{-i}).$$

That is, each agent i updates his beliefs according to his subjective prior but takes into account that the other agents have different subjective priors. Here, i's expected utility and incentive compatibility constraints are computed using μ_i.

In such a subjective setting, the designer has a subjective prior $\mu_d \in \Delta(\Theta)$, which together with a mechanism $\mathcal{M}$ induces the designer's ex-ante probability measure $\phi_d(\Theta \times \overline{Y} \times M)$. The designer's payoff is computed using ϕ_d. The revelation principle holds for an environment with subjective priors.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Additional definitions

A.1. Hierarchies of beliefs

Let $X_i^1 := \Theta_{-i}$ be the first-order domain of uncertainty of agent i and let $H_i^1 := \Delta(X_i^1)$ be the set of first-order beliefs of i. Inductively define the sets X_i^k, H_i^k as follows: Assume that the sets are defined for k. Let $X_i^{k+1} := X_i^k \times H_{-i}^k$ be the $(k+1)$-order domain of uncertainty of agent i and $H_i^{k+1} := \left\{(h_i^1, \ldots, h_i^k, h_i^{k+1}) \in H_i^k \times \Delta(X_i^{k+1}) : \operatorname{marg}_{X_i^k} h_i^{k+1} = h_i^k\right\}$ be the set of $(k+1)$-order beliefs of agent i. Notice that, if $(h_i^1, ..., h_i^{k+1}) \in H_i^{k+1}$, then $(h_i^1, ..., h_i^{\ell}) \in H_i^{\ell}$ for all $\ell \leq k$; that is, each $(h_i^1, ..., h_i^{k+1}) \in H_i^{k+1}$ is coherent. Then,

$$H_i := \left\{ (h_i^1, h_i^2, ...) \in \prod_{k=1}^{\infty} \Delta(X_i^k) : (h_i^1, h_i^2, ..., h_i^k) \in H_i^k \text{ for each } k \in \mathbb{N} \right\}$$

is the set of i's **collectively coherent hierarchies of beliefs.** Call $H := \prod_{i \in I} H_i$ the **belief structure**. Write $h_i = (h_i^1, h_i^2, ...) \in H_i$ for a hierarchy of beliefs and call h_i^k the k^{th}**-order belief**.

Brandenburger and Dekel (1993) constructs the canonical homeomorphism between the spaces H_i and $\Delta(\Theta_{-i} \times H_{-i})$.[19] For each $h_i = (h_i^1, h_i^2, ...) \in H_i$, there is unique probability measure $h_i^\infty \in \Delta(\Theta_{-i} \times H_{-i})$ so that $\text{marg}_{X_i^k} h_i^\infty = h_i^k$ for each $k \in \mathbb{N}$. Conversely, for each $h_i^\infty \in \Delta(\Theta_{-i} \times H_{-i})$ there is a unique $h_i = (h_i^1, h_i^2, ...) \in H_i$ so that for each $k \in \mathbb{N}$, $\text{marg}_{X_i^k} h_i^\infty = h_i^k$.

Fix an extensive form Γ and an associated profile of terminal belief mappings $\hat{\beta} = (\hat{\beta}_i)_{i \in I}$ with $\hat{\beta}_i : \hat{\mathcal{P}}_i \to \Delta(\hat{\mathcal{P}}_{-i})$. Set $\eta_i^1 : \hat{\mathcal{P}}_{-i} \to X_i^1$ so that $\eta_i^1(\hat{\rho}_{-i}) = \text{proj}_{\Theta_{-i}} \hat{\rho}_{-i}$ and note that η_i^1 is measurable. Assuming that the measurable maps η_i^k are defined, define $\eta_i^{k+1} : \hat{\mathcal{P}}_{-i} \to X_i^{k+1}$ so that $\eta_i^{k+1}(\hat{\rho}_{-i}) := (\eta_i^k(\hat{\rho}_{-i}), ((\underline{\eta_j^k} \circ \hat{\beta}_j)(\hat{\rho}_j))_{j \in I \setminus \{i\}})$, where $\underline{\eta_j^k} : \Delta(\hat{\mathcal{P}}_{-j}) \to \Delta(X_j^k)$ maps each measure in $\Delta(\hat{\mathcal{P}}_{-j})$ to the image measure under the mapping $\eta_j^k : \hat{\mathcal{P}}_{-j} \to X_j^k$. Note that, since η_j^k is measurable, $\underline{\eta_j^k}$ is also measurable. (See Theorem 15.14 in Aliprantis and Border (2006).) So, η_i^{k+1} is the composition of measurable functions and so, measurable.

Set $\hat{\delta}_i^k := \underline{\eta_i^k} \circ \hat{\beta}_i$. Note, that for each $\hat{\rho}_{-i}$, $\text{proj}_{X_i^k}(\eta_i^{k+1}(\hat{\rho}_{-i})) = \eta_i^k(\hat{\rho}_{-i})$. Thus, for each $\hat{\rho}_i \in \hat{\mathcal{P}}_i$, $\text{marg}_{X_i^k} \hat{\delta}_i^{k+1}(\hat{\rho}_i) = \hat{\delta}_i^k(\hat{\rho}_i)$. Write $\hat{\delta}_i : \hat{\mathcal{P}}_i \to H_i$ for $\hat{\delta}_i(\hat{\rho}_i) = (\hat{\delta}_i^1(\hat{\rho}_i), \hat{\delta}_i^2(\hat{\rho}_i), ...)$. Finally set $\hat{\delta} : \hat{\mathcal{P}} \to H$ as $\hat{\delta}(\hat{\rho}) = (\hat{\delta}_i(\hat{\rho}_i))_{i \in I}$.

A.2. Terminal information sets and expected utilities

Fix an extensive form Γ with associated strategy profile $S_0 \times S$. For each profile of behavior strategies $\sigma = (\sigma_i)_{i \in I}$, write $\sigma(\theta)(s)$ for $\prod_{j \in I} \sigma_j(\theta_j)(s_j)$.

Write $S(\hat{\rho}) := \left\{(s_0, s) \in S_0 \times S : (\text{proj}_\Theta(\hat{\rho}), \zeta(s_0, s)) \in \bigcap_{i \in I} \hat{\rho}_i\right\}$, for the set of strategy profiles that reach the terminal information sets $\hat{\rho} = (\hat{\rho}_i)_{i \in I}$. Define the ex-ante probability measure $\mathbb{P}(\rho \mid \sigma) \in \Delta(\hat{\mathcal{P}})$ by

$$\mathbb{P}(\hat{\rho} \mid \sigma) := \mu(\text{proj}_\Theta(\hat{\rho})) \sum_{(s_0,s) \in S(\hat{\rho})} \sigma(\text{proj}_\Theta(\hat{\rho}))(s) \cdot p_0(s_0).$$

So, $\mathbb{P}(\hat{\rho} \mid \sigma)$ is the ex-ante probability of reaching all information sets $\hat{\rho} = (\hat{\rho}_i)_{i \in I}$ given strategy profile σ. Similarly, define the ex-ante probability measure $\mathbb{P}(\cdot \mid s_i, \sigma_{-i}) \in \Delta(\hat{\mathcal{P}})$ by

$$\mathbb{P}(\hat{\rho} \mid s_i, \sigma_{-i}) := \mu(\text{proj}_\Theta(\hat{\rho})) \sum_{\substack{(s_0,s_{-i}) \in S_0 \times S_{-i}: \\ (s_0,s_i,s_{-i}) \in S(\hat{\rho})}} \sigma_{-i}(\text{proj}_\Theta(\hat{\rho}))(s_{-i}) \cdot p_0(s_0).$$

So, the value $\mathbb{P}(\hat{\rho} \mid, s_i, \sigma_{-i})$ is the ex-ante probability of reaching all the terminal information sets $\hat{\rho} = (\hat{\rho}_i)_{i \in I}$ given that agent i plays s_i and $-i$ play σ_{-i}.

In addition, define the interim probability measure $\mathbb{P}(\cdot \mid \theta, \sigma) \in \Delta(\hat{\mathcal{P}})$ by

$$\mathbb{P}(\hat{\rho} \mid \theta, \sigma) := \begin{cases} \displaystyle\sum_{(s_0,s) \in S(\hat{\rho})} \sigma(\theta)(s) \cdot p_0(s_0) & \text{if } \text{proj}_\Theta(\hat{\rho}) = \theta \\ 0 & \text{otherwise.} \end{cases}$$

So, the value $\mathbb{P}(\hat{\rho} \mid \theta_i, \sigma)$ is the probability of reaching $\hat{\rho} = (\hat{\rho}_i)_{i \in I}$ given that the type profile is θ and agents play σ. Define the interim probability measure $\mathbb{P}(\cdot \mid \theta_i, s_i, \sigma_{-i}) \in \Delta(\hat{\mathcal{P}})$ by

$$\mathbb{P}(\hat{\rho} \mid \theta_i, s_i, \sigma_{-i}) := \begin{cases} \beta_i(\theta_i)(\text{proj}_{\Theta_{-i}}(\hat{\rho})) \displaystyle\sum_{\substack{(s_0,s_{-i}) \in S_0 \times S_{-i}: \\ (s_0,s_i,s_{-i}) \in S(\hat{\rho})}} \sigma_{-i}(\text{proj}_{\Theta_{-i}}(\hat{\rho}))(s_{-i}) \cdot p_0(s_0) & \text{if } \text{proj}_{\Theta_i}(\hat{\rho}) = \theta_i \\ 0 & \text{otherwise.} \end{cases}$$

So, the value $\mathbb{P}(\hat{\rho} \mid \theta_i, s_i, \sigma_{-i})$ is the interim probability of reaching $\hat{\rho} = (\hat{\rho}_i)_{i \in I}$ given that agent i has type θ_i, plays s_i, and agents $-i$ play σ_{-i}.

Write

$$\mathbb{E}u_i(\theta_i, s_i, \mid \sigma_{-i}, \hat{\beta}) := \sum_{\hat{\rho} \in \hat{\mathcal{P}}} u_i(\text{proj}_\Theta(\hat{\rho}), \text{proj}_Y(\hat{\rho}), \hat{\delta}_i(\hat{\rho})) \cdot \mathbb{P}(\hat{\rho} \mid \theta_i, s_i, \sigma_{-i}),$$

where $\hat{\delta}_i$ is i's terminal hierarchy mapping induced by $\hat{\beta}$. So, $\hat{u}_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta})$ is the expected utility of type θ_i that plays s_i, given that the others play according to σ_{-i} and the terminal beliefs are $\hat{\beta}$. (Notice $\mathbb{P}(\hat{\rho} \mid \theta_i, s_i, \sigma_{-i}) = 0$ if $\text{proj}_{\Theta_i}(\hat{\rho}) \neq \theta_i$.) Similarly, write

$$\mathbb{E}\pi(\sigma, \hat{\beta}) := \sum_{\hat{\rho} \in \hat{\mathcal{P}}} \pi(\text{proj}_\Theta(\hat{\rho}), \text{proj}_Y(\hat{\rho}), \hat{\delta}(\hat{\rho})) \cdot \mathbb{P}(\hat{\rho} \mid \sigma),$$

for the designer's expected utility of a Bayesian equilibrium $(\sigma, \hat{\beta})$.

[19] The result in Brandenburger and Dekel (1993) uses a framework with a common space of uncertainty Θ_0, leading to a homeomorphism $H_i \to \Delta(\Theta_0 \times H_{-i})$. However, their proof extends (*mutatis mutandis*) to the present case.

Appendix B. Proofs of Section 2

Lemma B.1. *Fix a direct mechanism $\mathcal{M}:\Theta\to\Delta(Y)$ and consider the associated Bayesian game where each agent i only observes their material outcome y_i but not the report θ_{-i} nor y_{-i}. If the honest strategy profile is a Bayesian equilibrium, then either Ann has ex-ante expected utility different from 0 or the auctioneer has expected payoff different from 6.*

Proof. Bob only observes his material outcome $y_B=(0,0)$ and hence receives no information about Ann. Thus, regardless of Ann's report, $\mathbb{E}_A[\mathbb{E}_B[\theta_A]\,|\,h_A]=3<4$. A result, Ann's ex-ante expected psychological sub-utility is always 0.

Now consider the honest strategy profile and assume that Ann has ex-ante expected utility of 0. Then, the auctioneer can at most charge Ann's ex-ante valuation of the object plus Ann's ex-ante psychological sub-utility, otherwise Ann receives negative expected utility. Consequently, the auctioneer's expected payoff is capped at $3+0=3$. □

Lemma B.2. *Fix a direct mechanism $\mathcal{M}:\Theta\to\Delta(Y)$ and consider the associated Bayesian game where each agent i observes the reported profile θ, and the collective material outcome y. If the honest strategy profile is a Bayesian equilibrium, then either Ann has ex-ante expected utility different from 0 or the auctioneer has expected payoff different from 6.*

Proof. Under the honest strategy profile, Bob always fully learns Ann's type. As a result, a truthful report by Ann leads to $\mathbb{E}_A[\mathbb{E}_B[\theta_A]\,|\,h_A]=\theta_A$ for each $\theta_A\in\{0,6\}$. So, Ann's type $\theta_A=0$ receives psychological sub-utility 0 and Ann's type $\theta_A=6$ receives psychological sub-utility 4. Since types are equally likely, Ann's ex-ante psychological sub-utility equals 2.

Now assume that under the honest strategy profile Anna has ex-ante expected utility of 0. Then, ex-ante, the auctioneer can at most charge Ann's ex-ante valuation of the object plus Ann's ex-ante psychological sub-utility, otherwise Ann receives negative expected utility. Consequently, the auctioneer's expected payoff is capped at $3+2=5$. □

Appendix C. Proofs of Section 4

Lemma C.1. *Fix an extended direct mechanism $\mathcal{M}:\Theta\to\Delta(Y\times H)$. Let $\overline{Y}\times M$ the support of $\mathcal{M}$ and let $\widetilde{\Theta}\subseteq\Theta$ be non-empty. Then,*

$$\sum_{s_0\in(\overline{Y}\times M)^{\widetilde{\Theta}}}\prod_{\theta\in\widetilde{\Theta}}\mathcal{M}(\theta)(s_0(\theta))=1.$$

Proof. For each non-empty $\widetilde{\Theta}\subseteq\Theta$, there is some $k\geq 1$ such that $|\widetilde{\Theta}|=k$. We show that the result holds for any $\widetilde{\Theta}$ of cardinality k.

The proof is by induction on k. The base case $k=1$ follows from the fact, that for each $\theta\in\Theta$, $\sum_{(y,h)\in\overline{Y}\times M}\mathcal{M}(\theta)(y,h)=1$.

Suppose that the statement holds for any $\widetilde{\Theta}\subseteq\Theta$ with $|\widetilde{\Theta}|=k$. Fix some $\hat{\Theta}\subseteq\Theta$ with $|\hat{\Theta}|=k+1$ and write $\hat{\Theta}=\{\theta^1,...,\theta^k,\theta^{k+1}\}$ and consider $\widetilde{\Theta}=\{\theta^1,...,\theta^k\}$. Note that $(\overline{Y}\times M)^{\hat{\Theta}}=(\overline{Y}\times M)\times(\overline{Y}\times M)^{\widetilde{\Theta}}$. So, each function $s_0:\hat{\Theta}\to(\overline{Y}\times M)$ can be written as a singleton of $(\overline{Y}\times M)$ cross a function $\tilde{s}_0:\widetilde{\Theta}\to(\overline{Y}\times M)$. Thus,

$$\begin{aligned}
\sum_{s_0\in(\overline{Y}\times M)^{\hat{\Theta}}}\prod_{\ell=1}^{k+1}\mathcal{M}(\theta^\ell)(s_0(\theta^\ell)) &= \sum_{(y,h)\in\overline{Y}\times M}\sum_{\tilde{s}_0\in(\overline{Y}\times M)^{\widetilde{\Theta}}}\left(\mathcal{M}(\theta^{k+1})(y,h)\cdot\prod_{\ell=1}^{k}\mathcal{M}(\theta^\ell)(\tilde{s}_0(\theta^\ell))\right)\\
&= \sum_{(y,h)\in\overline{Y}\times M}\mathcal{M}(\theta^{k+1})(y,h)\cdot\left(\sum_{\tilde{s}_0\in(\overline{Y}\times M)^{\widetilde{\Theta}}}\prod_{\ell=1}^{k}\mathcal{M}(\theta^\ell)(\tilde{s}_0(\theta^\ell))\right)\\
&= \sum_{(y,h)\in\overline{Y}\times M}\mathcal{M}(\theta^{k+1})(y,h)\\
&= 1,
\end{aligned}$$

where the third equality follows from the induction hypothesis. □

Lemma C.2. *Fix a direct mechanism $\mathcal{M}$ with associated set of messages and material outcomes $\overline{Y}$ and M. Set $S_0=(\overline{Y}\times M)^{\Theta}$ and let $p_0:S_0\to\mathbb{R}$ be such that $p_0(s_0):=\prod_{\theta\in\Theta}\mathcal{M}(\theta)(s_0(\theta))$.*

(i) The mapping p_0 defines a probability measure on S_0.

(ii) For each $(\theta,y,h)\in\Theta\times\overline{Y}\times M$, $p_0(\{s_0\in S_0:s_0(\theta)=(y,h)\})=\mathcal{M}(\theta)(y,h)$.

Proof. Part (i) follows from $p_0(S_0)=\sum_{s_0\in S_0}\prod_{\theta\in\Theta}\mathcal{M}(\theta)(s_0(\theta))=1$. (See Lemma C.1.)

Fix $(\theta,y,h)\in\Theta\times Y\times M$. Write $\widetilde{\Theta}=\Theta\backslash\{\theta\}$. Notice that each $s_0\in S_0$ with $s_0(\theta)=(y,h)$ is associated with an unique $\tilde{s}_0\in(\overline{Y}\times M)^{\widetilde{\Theta}}$. Thus,

$$\begin{aligned} p_0\left(\left\{s_0 \in S_0 : s_0(\theta) = (y,h)\right\}\right) &= \sum_{s_0 \in S_0 : s_0(\theta)=(y,h)} \prod_{\theta' \in \Theta} \mathcal{M}(\theta')(s_0(\theta')) \\ &= \mathcal{M}(\theta)(y,h) \sum_{s_0 \in S_0 : s_0(\theta)=(y,h)} \prod_{\theta' \in \widetilde{\Theta}} \mathcal{M}(\theta')(s_0(\theta')) \\ &= \mathcal{M}(\theta)(y,h) \sum_{\widetilde{s_0} \in (\overline{Y} \times M)^{\widetilde{\Theta}}} \prod_{\theta' \in \widetilde{\Theta}} \mathcal{M}(\theta')(s_0(\theta')) \\ &= \mathcal{M}(\theta)(y,h) \cdot 1, \end{aligned}$$

where the last equality follows from Lemma C.1. This establishes (ii). □

Lemma C.3. *Fix an extended direct mechanism $\mathcal{M}$ with associated set of messages M. If $\mathcal{M}$ satisfies believability, then M is belief-closed.*

Proof. Let $\overline{Y}$ the set of feasible material outcomes associated to $\mathcal{M}$. Fix $h_i \in M_i$. It suffices to show $h_i^\infty(\Theta_{-i} \times M_{-i}) = 1$. Notice,

$$\begin{aligned} h_i^\infty(\Theta_{-i} \times M_{-i}) \cdot \operatorname{marg}_{M_i} \phi(h_i) &= \sum_{(\theta_{-i},h_{-i}) \in \Theta_{-i} \times M_{-i}} \sum_{(\theta_i,y_i) \in \Theta_i \times Y} h_i^\infty(\theta_{-i},h_{-i}) \cdot \operatorname{marg}_{M_i} \phi(\theta_i,y_i,h_i) \\ &= \sum_{(\theta_{-i},h_{-i}) \in \Theta_{-i} \times M_{-i}} \sum_{(\theta_i,y_i) \in \Theta_i \times Y} \operatorname{marg}_{\Theta \times Y_i \times M} \phi(\theta_i,\theta_{-i},y_i,h_i,h_{-i}) \\ &= \operatorname{marg}_{M_i} \phi(h_i), \end{aligned}$$

where the second equality follows from the fact that $\mathcal{M}$ satisfies believability. Moreover, since $h_i \in M_i$, there is some $(\theta, y, h_{-i}) \in \Theta \times \overline{Y} \times M_{-i}$ such that $\operatorname{marg}_{M_i} \phi(h_i) \geq \phi(\theta,y,h_i,h_{-i}) = \mathcal{M}(\theta)(y,h_i,h_{-i}) > 0$. Therefore, $h_i^\infty(\Theta_{-i} \times M_{-i}) = 1$, as desired. □

Lemma C.4. *Fix $\hat{\rho}_i \in \hat{\mathcal{P}}_i$ and let $\hat{\delta}_i(\hat{\rho}_i) = h_i$.*
(i) For each $(\theta_{-i},h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$, $h_i^\infty(\theta_{-i},h_{-i}) = \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i},h_{-i}])$.
(ii) $\operatorname{Supp} h_i^\infty \subseteq \Theta_i \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$.

Proof. Set $\hat{\delta}_i(\hat{\rho}_i) = h_i = (h_i^1, h_i^2, \ldots)$. To show (i), fix $(\theta_{-i},h_{-i}) \in \Theta_{-i} \times H_{-i}$ with $h_{-i} = (h_{-i}^1, h_{-i}^2, \ldots)$. For each $k \in \mathbb{N}$, write $B^k := \{(\theta_{-i}, h_{-i}^1 \ldots, h_{-i}^k)\} \times \prod_{\ell=k+1}^\infty \prod_{j \in I \setminus \{i\}} \Delta(X_j^\ell)$. Note, each B^k is a measurable subset of $\Theta_{-i} \times H_{-i}$. Moreover, $(B^k)_{k \in \mathbb{N}}$ is a sequence of decreasing sets with $\bigcap_{k \in \mathbb{N}} B^k = \{(\theta_{-i},h_{-i})\}$. Thus,

$$h_i^\infty(\theta_i,h_{-i}) = h_i^\infty\left(\textstyle\bigcap_{k \in \mathbb{N}} B^k\right) = \lim_{k \to \infty} h_i^\infty(B^k) = \lim_{k \to \infty} h_i^{k+1}(\theta_{-i}, h_{-i}^1, \ldots, h_{-i}^k), \tag{2}$$

where the last equality follows from the fact that $\operatorname{marg}_{X_i^{k+1}} h_i^\infty = h_i^{k+1}$.

Now, since $\hat{\delta}_i(\hat{\rho}_i) = (h_i^1, h_i^2, \ldots)$ and η_i^{k+1} is measurable,

$$h_i^{k+1}(\theta_{-i}, h_{-i}^1, \ldots, h_{-i}^k) = ((\eta_i^{k+1} \circ \hat{\beta}_i)(\hat{\rho}_i))(\theta_{-i}, h_{-i}^1, \ldots, h_{-i}^k) = \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, \ldots, h_{-i}^k]), \tag{3}$$

where $\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, \ldots h_{-i}^k] := (\eta_i^{k+1})^{-1}(\theta_{-i}, h_{-i}^1, \ldots h_{-i}^k)$. Notice that $(\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, \ldots h_{-i}^k])_{k \in \mathbb{N}}$ is a decreasing sequence of measurable sets such that $\hat{\mathcal{P}}_{-i}[\theta_{-i},h_{-i}] = \bigcap_{k \in \mathbb{N}} \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, \ldots h_{-i}^k]$. Consequently,

$$\begin{aligned} \lim_{k \to \infty} \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, \ldots h_{-i}^k]) &= \hat{\beta}_i(\hat{\rho}_i)\left(\textstyle\bigcap_{k \in \mathbb{N}} \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, \ldots h_{-i}^k]\right) \\ &= \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i},h_{-i}]). \end{aligned} \tag{4}$$

So,

$$\begin{aligned} h_i^\infty(\theta_{-i},h_{-i}) &= \lim_{k \to \infty} h_i^{k+1}(\theta_{-i}, h_{-i}^1, \ldots, h_{-i}^k) \\ &= \lim_{k \to \infty} \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, \ldots h_{-i}^k]) \\ &= \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i},h_{-i}]), \end{aligned} \tag{5}$$

where the first equality follows from Equation (2), the second from Equation (3), and the third from Equation (4). This establishes (i).

To show (ii), notice that $\Theta_i \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$ is finite and, thus, closed. In addition, note that

$$\bigcup_{(\theta_{-i},h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})} \hat{\mathcal{P}}_{-i}[\theta_{-i},h_{-i}] = \hat{\mathcal{P}}_{-i}.$$

So, by Equation (5), $h_i^\infty(\Theta_i \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})) = \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}) = 1$. This establishes (ii). □

Write $\tau_i : \Theta_i \times S_0 \times S \to \hat{\mathcal{P}}_i$ for the mapping such that, for each type profile $\theta_{-i} \in \Theta_{-i}$, $(\theta_i, \theta_{-i}, \zeta(s_0, s)) \in \tau_i(\theta_i, s_0, s)$. That is, $\tau_i(\theta_i, s_0, s) \in \hat{\mathcal{P}}_i$ specifies the terminal information set of agent i, given that i is of type θ_i and the agents play (s_0, s). Write $\tau : \Theta \times S_0 \times S \to \hat{\mathcal{P}}$ so that $\tau(\theta, s_0, s) = (\tau_i(\theta_i, s_0, s))_{i \in I}$.

Lemma C.5. *Fix a consistent profile $(\sigma, \hat{\beta})$, types $\theta_i, \theta_i' \in \Theta_{-i}$, and $(s_0, s) \in S_0 \times S$ so that $p_0(s_0) \cdot \sigma_{-i}(\theta_{-i}')(s_{-i}) > 0$ for some $\theta_{-i}' \in \Theta_{-i}$. Write $\hat{\delta}_i(\tau_i(\theta_i, s_0, s)) = h_i$ and $\hat{\delta}_i(\tau_i(\theta_i', s_0, s)) = h_i'$ for the associated hierarchies of beliefs. Then, there is some $c > 0$ so that, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$, $h_i^{\infty}(\theta_{-i}, h_{-i}) = c \, \frac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \, h_i'^{\infty}(\theta_{-i}, h_{-i})$.*

Proof. Write $\tau_i(\theta_i, s_0, s) = \hat{\rho}_i$, $\tau_i(\theta_i', s_0, s) = \hat{\rho}_i'$; so $\hat{\delta}(\hat{\rho}_i) = h_i$ and $\hat{\delta}(\hat{\rho}_i') = h_i'$. By Lemma C.4, $\mathrm{Supp}(h^{\infty})$ and $\mathrm{Supp}(h'^{\infty})$ are subsets of $\Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$. Moreover, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$,

(i) $h_i^{\infty}(\theta_{-i}, h_{-i}) = \sum_{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]} \hat{\beta}(\hat{\rho}_i)(\hat{\rho}_{-i})$, and

(ii) $h_i'^{\infty}(\theta_{-i}, h_{-i}) = \sum_{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]} \hat{\beta}(\hat{\rho}_i')(\hat{\rho}_{-i})$.

Thus, it suffices to show that there is a $c > 0$ so that, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$ and each $\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]$,

$$\hat{\beta}_i(\hat{\rho}_i, \hat{\rho}_{-i}) = c \, \tfrac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \, \hat{\beta}_i(\hat{\rho}_i', \hat{\rho}_{-i}). \tag{6}$$

Write $x := \sum_{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}} \mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid s_i, \sigma_{-i})$ and $x' := \sum_{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}} \mathbb{P}(\hat{\rho}_i', \hat{\rho}_{-i} \mid s_i, \sigma_{-i})$. We first show that $x, x' > 0$. Then, we show that Equation (6) holds for $c = x'/x$.

Let $\theta_{-i}' \in \Theta_{-i}$ be such that $p_0(s_0)\, \sigma_{-i}(\theta_{-i}')(s_{-i}) > 0$ and fix $\hat{h}_{-i}' = (\tau_j(\theta_j', s_0, s))_{j \in I \setminus \{i\}}$. To show $x > 0$, it suffices to show $\mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i}' \mid s_i, \sigma_{-i}) > 0$. Observe that

$$\begin{aligned}
\mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i}' \mid s_i, \sigma_{-i}) &= \mu(\theta_i, \theta_{-i}') \sum_{\substack{(s_0', s_{-i}') \in S_0 \times S_{-i}: \\ (s_0', s_i, s_{-i}') \in S(\hat{\rho}_i, \hat{\rho}_{-i}')}} p_0(s_0') \cdot \sigma_{-i}(\theta_{-i}')(s_{-i}') \\
&\geq \mu(\theta_i, \theta_{-i}') \cdot p_0(s_0) \cdot \sigma_{-i}(\theta_{-i}')(s_{-i}) \\
&> 0,
\end{aligned}$$

where the first equality follows from the definition of $\mathbb{P}(\cdot \mid \cdot)$, the first inequality from the fact that $(s_0, s_i, s_{-i}) \in S(\hat{\rho}_i, \hat{\rho}_{-i})$, and the last inequality by assumption. This establishes $x > 0$; an analogous argument establishes $x' > 0$.

Set $c = x'/x$, note that $c > 0$. Fix $\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]$. To show that Equation (6) holds, it suffices to show the following:

(a) $x \cdot \hat{\beta}_i(\hat{\rho}_i, \hat{\rho}_{-i}) = \mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid s_i, \sigma_{-i})$,

(b) $x' \cdot \hat{\beta}_i(\hat{\rho}_i', \hat{\rho}_{-i}) = \mathbb{P}(\hat{\rho}_i', \hat{\rho}_{-i} \mid s_i, \sigma_{-i})$, and

(c) $\mathbb{P}((\hat{\rho}_i, \hat{\rho}_{-i}) \mid s_i, \sigma_{-i}) = \frac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \mathbb{P}(\hat{\rho}_i', \hat{\rho}_{-i} \mid s_i, \sigma_{-i})$.

If so, then

$$\hat{\beta}_i(\hat{\rho}_i, \hat{\rho}_{-i}) = \tfrac{1}{x} \mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid s_i, \sigma_{-i}) = \tfrac{1}{x} \tfrac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \mathbb{P}(\hat{\rho}_i', \hat{\rho}_{-i} \mid s_i, \sigma_{-i}) = \tfrac{x'}{x} \tfrac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \hat{\beta}_i(\hat{\rho}_i', \hat{\rho}_{-i}),$$

as desired.

Equalities (a) and (b) follow from consistency of $(\sigma, \hat{\beta})$. To show (c), recall that $\hat{\rho}_i = \tau_i(\theta_i, s_0, s)$ and $\hat{\rho}_i' = \tau_i(\theta_i', s_0, s)$. So, $\hat{\rho}_i$ and $\hat{\rho}_i'$ are associated to the same partition element $\rho_i \in \mathcal{P}_i$. Thus,

$$\begin{aligned}
S(\hat{\rho}_i, \hat{\rho}_{-i}) &= \left\{(s_0', s') \in S_0 \times S : ((\theta_i, \theta_{-i}), \zeta(s_0', s')) \in \bigcap (\hat{\rho}_i, \hat{\rho}_{-i})\right\} \\
&= \left\{(s_0', s') \in S_0 \times S : ((\theta_i', \theta_{-i}), \zeta(s_0', s')) \in \bigcap (\hat{\rho}_i', \hat{\rho}_{-i})\right\} \\
&= S(\hat{\rho}_i', \hat{\rho}_{-i}).
\end{aligned} \tag{7}$$

That is, $(\hat{\rho}_i, \hat{\rho}_{-i})$ and $(\hat{\rho}_i', \hat{\rho}_{-i})$ are reached by the same subset of $S_0 \times S$. Therefore,

$$\begin{aligned}
\mathbb{P}((\hat{\rho}_i, \hat{\rho}_{-i}) \mid s_i, \sigma_{-i}) &= \mu(\theta_i, \theta_{-i}) \sum_{\substack{(s_0', s_{-i}') \in S_0 \times S_{-i}: \\ (s_0', s_i, s_{-i}') \in S(\hat{\rho}_i, \hat{\rho}_{-i})}} p_0(s_0) \cdot \sigma_{-i}(\theta_{-i})(s_{-i}) \\
&= \frac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \, \mu(\theta_i', \theta_{-i}) \sum_{\substack{(s_0', s_{-i}') \in S_0 \times S_{-i}: \\ (s_0', s_i, s_{-i}') \in S(\hat{\rho}_i', \hat{\rho}_{-i})}} p_0(s_0) \cdot \sigma_{-i}(\theta_{-i})(s_{-i}) \\
&= \frac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \mathbb{P}((\hat{\rho}_i', \hat{\rho}_{-i}) \mid s_i, \sigma_{-i}),
\end{aligned}$$

where the second equality follows by Equation (7). □

Corollary C.1. *Fix a consistent profile $(\sigma, \hat{\beta})$, types $\theta_i, \theta_i' \in \Theta_{-i}$, and $(s_0, s) \in S_0 \times S$ so that $p_0(s_0) \cdot \sigma_{-i}(\theta_{-i}')(s_{-i}) > 0$ for some $\theta_{-i}' \in \Theta_{-i}$. If the prior μ is independent, then*

$$\hat{\delta}_i(\tau_i(\theta_i, s_0, s)) = \hat{\delta}_i(\tau_i(\theta_i', s_0, s)).$$

Proof. Write $\hat{\delta}_i(\tau_i(\theta_i, s_0, s)) = h_i$ and $\hat{\delta}_i(\tau_i(\theta_i', s_0, s)) = h_i'$ for the associated hierarchies of beliefs. We show $h_i = h_i'$. From Lemma C.5, there is some $c > 0$ so that, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$,

$$h_i^{\infty}(\theta_{-i}, h_{-i}) = c \ \frac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \ h_i'^{\infty}(\theta_{-i}, h_{-i}).$$

Now write $\hat{c} := c \ \frac{\mu(\{\theta_i\} \times \Theta_{-i})}{\mu(\{\theta_i'\} \times \Theta_{-i})}$. Note, by independence, $c \ \frac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} = c \ \frac{\mu(\{\theta_i\} \times \Theta_{-i})}{\mu(\{\theta_i'\} \times \Theta_{-i})} = \hat{c}$. Thus, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$, $h_i^{\infty}(\theta_{-i}, h_{-i}) = \hat{c} \cdot h_i'^{\infty}(\theta_{-i}, h_{-i})$. Since h_i^{∞} and $h_i'^{\infty}$ are both probability measures,

$$\sum_{(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})} h_i^{\infty}(\theta_{-i}, h_{-i}) = \sum_{(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})} h_i'^{\infty}(\theta_{-i}, h_{-i}) = 1.$$

So, $\hat{c} = 1$ and $h_i^{\infty} = h'^{\infty}_i$. □

Write $\hat{\mathcal{P}}_i^+$ for i's collection of terminal information sets that are reached on the honest equilibrium path. Write $R_{-i} \subseteq \hat{\mathcal{P}}_{-i}$ for the terminal information sets of $-i$ consistent with a honest report, i.e., the profiles $[\theta_{-i}, \theta_{-i}', y_{-i}, h_{-i}] \in \hat{\mathcal{P}}_{-i}$ such that $\theta_i' = \theta_i$. Note that $\hat{\mathcal{P}}_{-i}^+ \subseteq R_{-i}$. Write $R_{-i}[\tilde{\theta}_{-i}]$ for the terminal information sets in R_{-i} consistent with types $\tilde{\theta}_{-i}$, i.e., the profiles $[\theta_{-i}, \theta_{-i}, y_{-i}, h_{-i}] \in R_{-i}$ such that $\theta_{-i} = \tilde{\theta}_{-i}$. For each $k \in \mathbb{N}$, write $R_{-i}[\tilde{\theta}_{-i}, \tilde{h}_{-i}^1, ..., \tilde{h}_{-i}^k]$ for the terminal information sets in R_{-i} consistent with types $\tilde{\theta}_{-i}$ and messages $(\tilde{h}_{-i}^1, ..., \tilde{h}_{-i}^k)$, i.e., the profiles $[\theta_{-i}, \theta_{-i}, y_{-i}, h_{-i}] \in R_{-i}$ such that $\theta_{-i} = \tilde{\theta}_{-i}$ and $\text{proj}_{H_{-i}^k}(h_{-i}) = (\tilde{h}_{-i}^1, ..., \tilde{h}_{-i}^k)$.

Proof of Lemma 4.1. Fix a profile of honest beliefs $\hat{\beta}^*$ and let $\hat{\delta}^*$ be the terminal hierarchies of beliefs it induces. The proof is divided in three steps. The first computes $\hat{\beta}_i^*(\hat{\rho}_i)$ for each $\hat{\rho}_i \in \hat{\mathcal{P}}_i^+$. The second uses the first step to show part (i). The third uses part (i) to show part (ii).

Step 1. Fix $\hat{\rho}_i = [\theta_i, \theta_i, y_i, h_i] \in \hat{\mathcal{P}}_i^+$, where $h_i = (h_i^1, h_i^2, ...)$. We show that

(a) $\text{Supp}\, \hat{\beta}_i^*(\hat{\rho}_i) \subseteq \hat{\mathcal{P}}_{-i}^+$,

(b) for each $\theta_{-i} \in \Theta_{-i}$, $\hat{\beta}_i^*(\hat{\rho}_i)(R_{-i}[\theta_{-i}]) = h_i^1(\theta_{-i})$,

(c) for each $(\theta_{-i}, h_{-i}^1, ..., h_{-i}^k) \in \Theta_{-i} \times H_{-i}^k$, $\hat{\beta}_i^*(\hat{\rho}_i)(R_{-i}[\theta_{-i}, h_{-i}^1, ..., h_{-i}^k]) = h_i^{k+1}(\theta_{-i}, h_{-i}^1, ..., h_{-i}^k)$.

To show (a)-(c), we first compute each $\hat{\beta}_i^*(\hat{\rho}_i)(\hat{\rho}_{-i})$. Fix $\hat{\rho}_{-i} = [\theta_{-i}, \theta_{-i}, y_{-i}, h_{-i}] \in \hat{\mathcal{P}}_{-i}$. Notice that consistency of $\hat{\beta}^*$ and σ^* states that if $s_i^* = \theta_i$, then

$$\hat{\beta}_i^*(\hat{\rho}_i)(\hat{\rho}_{-i}) \sum_{\hat{\rho}_{-i}' \in \hat{\mathcal{P}}_{-i}} \mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i}' \mid s_i^*, \sigma_{-i}^*) = \mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid s_i^*, \sigma_{-i}^*). \tag{8}$$

Notice, if $\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i} \backslash R_{-i}$, then $\mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid s_i^*, \sigma_{-i}^*) = 0$, since σ_{-i}^* does not allow false reports. Thus, we focus on $\hat{\rho}_{-i} \in R_{-i}$. Note, if $s_i^* = \theta_i$, then

$$\mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} \mid s_i^*, \sigma_{-i}^*) = \mu(\theta_i, \theta_{-i}) \cdot \mathcal{M}(\theta_i, \theta_{-i})(y_i, y_{-i}, h_i, h_{-i}) = \phi(\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i}), \tag{9}$$

where the equalities follow from definition of $\mathcal{M}$ and ϕ. So,

$$\sum_{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}} \mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} | s_i^*, \sigma_{-i}^*) = \sum_{\hat{\rho}_{-i} \in R_{-i}} \mathbb{P}((\hat{\rho}_i, \hat{\rho}_{-i}) | s_i^*, \sigma_{-i}^*) = \text{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i), \tag{10}$$

where the first equality follows from the fact that $\mathbb{P}(\hat{\rho}_i, \hat{\rho}_{-i} | s_i^*, \sigma_{-i}^*) = 0$ if $\hat{\rho}_{-i} \notin R_{-i}$ and the second from Equation (9). In addition, recall that $\hat{\rho}_i \in \hat{\mathcal{P}}_i^+$ means $\text{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i) > 0$. Therefore, Equations (8), (9), and (10) imply that for each $[\theta_{-i}, \theta_{-i}, y_{-i}, h_{-i}] \in R_{-i}$,

$$\hat{\beta}_i^*(\hat{\rho}_i)([\theta_{-i}, \theta_{-i}, y_{-i}, h_{-i}]) = \frac{\phi(\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i})}{\text{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i)}. \tag{11}$$

We use Equation (11) to show (a)-(c). To show (a), fix $\hat{\rho}_{-i} = [\theta_{-i}, \theta_{-i}, y_{-i}, h_{-i}] \in H_{-i}$ such that $\hat{\beta}_i^*(\hat{\rho}_i)(\hat{\rho}_{-i}) > 0$. By Equation (8), it follows that $\hat{\rho}_{-i} \in R_{-i}$. So, by Equation (11), $\phi(\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i}) > 0$. Therefore, $\text{Supp}\, \hat{\beta}_i^*(\hat{\rho}_i) \subseteq \hat{\mathcal{P}}_{-i}^+$, establishing (a).

To show (b), fix $\theta_{-i} \in \Theta_{-i}$. Note that

$$\hat{\beta}_i^*(\hat{\rho}_i)(R_{-i}[\theta_{-i}]) = \sum_{h_{-i}' \in M_{-i}} \sum_{y_{-i}' \in Y_{-i}} \hat{\beta}_i^*(\hat{\rho}_i)([\theta_{-i}, \theta_{-i}, y_{-i}', h_{-i}'])$$

$$= \sum_{h'_{-i} \in M_{-i}} \sum_{y'_{-i} \in Y_{-i}} \frac{\phi(\theta_i, \theta_{-i}, y_i, y'_{-i}, h_i, h'_{-i})}{\operatorname{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i)}$$

$$= \sum_{h'_{-i} \in M_{-i}} h_i^{\infty}(\theta_{-i}, h'_{-i})$$

$$= h_i^1(\theta_{-i}),$$

where the first equality follows from definition of $R_{-i}[\theta_{-i}]$, the second follows from Equation (11), the third by believability of $\mathcal{M}$, and the last from the fact that $\operatorname{marg}_{X_i^1} h_i^{\infty} = h_i^1$. This establishes (b).

To show (c), fix $(\theta_{-i}, h_{-i}^1, ..., h_{-i}^k) \in \Theta_{-i} \times H_{-i}^k$ and write $M_{-i}[h_{-i}^1, ..., h_{-i}^k] = \{h'_i \in M_{-i} : \operatorname{proj}_{H_i^k} h'_i := (h_{-i}^1, ..., h_{-i}^k)\}$, for the set of messages for $-i$ that is associated with $(h_{-i}^1, ..., h_{-i}^k)$. Note that

$$\hat{\beta}_i^*(\hat{\rho}_i)(R_{-i}[\theta_{-i}, h_{-i}^1, ..., h_{-i}^k]) = \sum_{h'_{-i} \in M_{-i}[h_{-i}^1, ..., h_{-i}^k]} \sum_{y'_{-i} \in Y_{-i}} \hat{\beta}_i^*(\hat{\rho}_i)([\theta_{-i}, \theta_{-i}, y'_{-i}, h'_{-i}])$$

$$= \sum_{h'_{-i} \in M_{-i}[h_{-i}^1, ..., h_{-i}^k]} \sum_{y'_{-i} \in Y_{-i}} \frac{\phi(\theta_i, \theta_{-i}, y_i, y'_{-i}, h_i, h'_{-i})}{\operatorname{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i)}$$

$$= \sum_{h'_{-i} \in M_{-i}[h_{-i}^1, ..., h_{-i}^k]} h_i^{\infty}(\theta_{-i}, h'_{-i})$$

$$= h_i^{k+1}(\theta_{-i}, h_{-i}^1, ..., h_{-i}^k),$$

where the first equality follows from the definition of $R_{-i}[\theta_{-i}, h_{-i}^1, ..., h_{-i}^k]$, the second from Equation (11), the third from believability of $\mathcal{M}$, and the last from the fact that $\operatorname{marg}_{X_i^{k+1}} h_i^{\infty} = h_i^{k+1}$. This establishes (c).

Step 2. We show part (i). It suffices to show that if $[\theta_i, \theta_i, y_i, h_i] \in \hat{\mathcal{P}}_i^+$, then $\hat{\delta}_i^{*k}([\theta_i, \theta_i, y_i, h_i]) = h_i^k$ for each $k \in \mathbb{N}$. The proof is by induction on k.

First take $k = 1$. Write $\hat{\mathcal{P}}_{-i}[\theta_{-i}] = \{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i} : \eta_i^1(\hat{\rho}_{-i}) = \theta_{-i}\}$ for the terminal information sets of $-i$ consistent with θ_{-i}. Fix $\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}^+$ and note that $\hat{\rho}_{-i} \in R_{-i}[\theta_{-i}]$ if and only if $\eta_i^1(\hat{\rho}_{-i}) = \theta_{-i}$. Therefore,

$$R_{-i}[\theta_{-i}] \cap \hat{\mathcal{P}}_{-i}^+ = \hat{\mathcal{P}}_{-i}[\theta_{-i}] \cap \hat{\mathcal{P}}_{-i}^+. \tag{12}$$

Thus,

$$\begin{aligned}
\hat{\delta}_i^{1*}(\hat{\rho}_i)(\theta_{-i}) &= (\underline{\eta}_i^1 \circ \hat{\beta}_i^*)(\hat{\rho}_i)(\theta_{-i}) \\
&= \hat{\beta}_i^*(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}]) \\
&= \hat{\beta}_i^*(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}] \cap \hat{\mathcal{P}}_{-i}^+) \\
&= \hat{\beta}_i^*(\hat{\rho}_i)(R_{-i}[\theta_{-i}] \cap \hat{\mathcal{P}}_{-i}^+) \\
&= \hat{\beta}_i^*(\hat{\rho}_i)(R_{-i}[\theta_{-i}]) \\
&= h_i^1(\theta_{-i}),
\end{aligned} \tag{13}$$

where the third and the fifth equalities follow from (a), the fourth from Equation (12) and the last one from (b).

Now, assume the claim is true for each $\ell \in \{1, ..., k\}$. We show that the claim is true for $k+1$. Write $\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, ..., h_{-i}^k] = \{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i} : \eta_i^{k+1}(\hat{\rho}_{-i}) = (\theta_{-i}, h_{-i}^1, ..., h_{-i}^k)\}$, for the set of terminal information sets of $-i$ consistent with type θ_{-i} and messages $(h_{-i}^1, ..., h_{-i}^k)$. Fix $\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}^+$. Note that, $\hat{\rho}_{-i} \in R_{-i}[\theta_{-i}, h_{-i}^1, ..., h_{-i}^k]$ if and only if

$$\eta_i^{k+1}(\hat{\rho}_{-i}) = (\operatorname{proj}_{\Theta_{-i}}(\hat{\rho}_{-i}), \hat{\delta}_{-i}^1(\hat{\rho}_{-i}), ..., \hat{\delta}_{-i}^k(\hat{\rho}_{-i})) = (\theta_{-i}, h_{-i}^1, ..., h_{-i}^k),$$

where the second equality follows by the induction hypothesis. Therefore,

$$R_{-i}[\theta_{-i}, h_{-i}^1, ..., h_{-i}^k] \cap \hat{\mathcal{P}}_i^+ = \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}^1, ..., h_{-i}^k] \cap \hat{\mathcal{P}}_i^+. \tag{14}$$

By using an analogous argument as in Equation (13), for each $(\theta_{-i}, h_{-i}^1, ..., h_{-i}^k) \in \Theta_{-i} \times H_{-i}^k$, $\hat{\delta}_i^{k+1*}(\hat{\rho}_i)(\theta_{-i}, h_{-i}^1, ..., h_{-i}^k) = h_i^{k+1}(\theta_{-i}, h_{-i}^1, ..., h_{-i}^k)$.

Step 3. We show part (ii) Assume that types are independent. Fix $[\theta_i, \theta_i, y_i, h_i] \in \hat{\mathcal{P}}_i^+$ and $\theta'_i \in \Theta_i$. Then, $\hat{\delta}_i^*([\theta_i, \theta'_i, y_i, h_i]) = \hat{\delta}_i^*([\theta_i, \theta_i, y_i, h_i]) = h_i$, where the first equality follows from Corollary C.1 and the second from Step 2. □

Lemma C.6. *Let $\mathcal{M}$ extended direct mechanism that satisfies BLV and BIC. Then, each honest profile $(\sigma^*, \hat{\beta}^*)$ constitutes a Bayesian equilibrium of $(\Gamma(\mathcal{M}), \mu)$.*

Proof. Let (σ^*, β^*) be a honest profile. By BIC, there exist honest terminal beliefs $\hat{\beta}^{**}$ such that, for each $\theta_i, \theta_i' \in \Theta_i$, $\mathbb{E}u_i(\theta_i, \theta_i \mid \sigma^*_{-i}, \hat{\beta}^{**}) \geq \mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma^*_{-i}, \hat{\beta}^{**})$. It is sufficient to show

$$\mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma^*_{-i}, \hat{\beta}^*) = \mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma^*_{-i}, \hat{\beta}^{**}). \tag{15}$$

From this, it follows that for each $\theta_i, \theta_i' \in \Theta_i$, $\mathbb{E}u_i(\theta_i, \theta_i \mid \sigma_i^*, \hat{\beta}_i^*) \geq \mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma_i^*, \hat{\beta}_i^*)$. So (σ^*, β^*) is a Bayesian equilibrium, as desired.

Let $\hat{\delta}^*$ (resp. $\hat{\delta}^{**}$) be the terminal hierarchies of beliefs induced by $\hat{\beta}^*$ (resp. $\hat{\beta}^{**}$). To show Equation (15), it suffices to show that $\hat{\delta}_i^*$ and $\hat{\delta}_i^{**}$ are identical at each terminal information set consistent with $\mathcal{M}(\theta_i', \theta_{-i})(y,h) > 0$.

Fix a profile $(\theta_i, \theta_{-i}) \in \Theta$, report θ_i', and outcome $(y,h) \in \overline{Y} \times M$ with $\mathcal{M}(\theta_i', \theta_{-i})(y,h) > 0$. Write $\hat{\delta}_i^*([\theta_i, \theta_i', y_i, h_i]) = h_i^*$ and $\hat{\delta}_i^{**}([\theta_i, \theta_i', y_i, h_i]) = h_i^{**}$. We will show that h_i^* is a multiple of h_i^{**}. Thus, since h_i^*, h_i^{**} are sequences of probability measures, it follows that $h_i^* = h_i^{**}$.

We show that h_i^* is a multiple of h_i^{**}. Note that $\mathbb{P}[[\theta_i', \theta_i', y_i, h_i] \mid \sigma^*] = \mathcal{M}(\theta_i', \theta_i)(y,h) > 0$, so $[\theta_i', \theta_i', y_i, h_i] \in \hat{\mathcal{P}}_i^+$. Then, since $\mathcal{M}$ satisfies BLV, $\hat{\delta}^*([\theta_i', \theta_i', y_i, h_i]) = h_i$ and $\hat{\delta}^{**}([\theta_i', \theta_i', y_i, h_i]) = h_i$. (See Lemma 4.1.)

Recall that $\mathcal{M}(\theta_i', \theta_{-i})(y,h) > 0$ means that chance selects (y,h) with positive probability after the report (θ_i', θ_{-i}). Thus, there is some $s_0 \in S_0$ such that $s_0(\theta_i', \theta_{-i}) = (y,h)$ with $p_0(s_0) > 0$. Fix such $s_0 \in S_0$ and notice that for reports $s = (\theta_i', \theta_{-i})$, $\tau_i(\theta_i', s_0, s) = [\theta_i', \theta_i', y_i, h_i]$ and $\tau_i(\theta_i, s_0, s) = [\theta_i, \theta_i', y_i, h_i]$. So,

(i) $\hat{\delta}_i^*(\tau_i(\theta_i', s_0, s)) = \hat{\delta}_i^*([\theta_i', \theta_i', y_i, h_i]) = h_i$, and
(ii) $\hat{\delta}_i^*(\tau_i(\theta_i, s_0, s)) = \hat{\delta}_i^*([\theta_i, \theta_i', y_i, h_i]) = h_i^*$.

Thus, by Lemma C.5, there is $c^* > 0$ so that, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times M_{-i}$,

$$h_i^{*\infty}(\theta_{-i}, h_{-i}) = c^* \, \frac{\mu(\theta_i', \theta_{-i})}{\mu(\theta_i, \theta_{-i})} \, h_i^{\infty}(\theta_{-i}, h_{-i}). \tag{16}$$

Similarly, write $\hat{\delta}_i^{**}([\theta_i, \theta_i', y_i, h_i]) = h_i^{**}$ and notice that

(iii) $\hat{\delta}_i^{**}(\tau_i(\theta_i', s_0, s)) = \hat{\delta}_i^{**}([\theta_i', \theta_i', y_i, h_i]) = h_i$, and
(iv) $\hat{\delta}_i^{**}(\tau_i(\theta_i, s_0, s)) = \hat{\delta}_i^{**}([\theta_i, \theta_i', y_i, h_i]) = h_i^{**}$.

Thus, by Lemma C.5, there is $c^{**} > 0$ so that, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times M_{-i}$,

$$h_i^{**\infty}(\theta_{-i}, h_{-i}) = c^{**} \, \frac{\mu(\theta_i', \theta_{-i})}{\mu(\theta_i, \theta_{-i})} \, h_i^{\infty}(\theta_{-i}, h_{-i}). \tag{17}$$

By Lemma C.4, $\operatorname{Supp} h_i^{*\infty}$ and $\operatorname{Supp} h_i^{**\infty}$ are subsets of $\Theta_{-i} \times M_{-i}$. Summing up across Equations (16) and (17), $h_i^{*\infty} = \frac{c^*}{c^{**}} \cdot h_i^{**\infty}$. So $h_i^* = \frac{c^*}{c^{**}} \cdot h_i^{**}$, as desired. □

Lemma C.7. *Let $\mathcal{M}$ be an extended direct mechanism satisfying BLV. Then, for each honest profile $(\sigma^*, \hat{\beta}^*)$, the following hold:*

(i) $\mathbb{E}\pi(\sigma^* \mid \hat{\beta}^*) = \sum_{(\theta,y,h) \in \Theta \times \overline{Y} \times M} \pi(\theta, y, h) \cdot \phi(\theta, y, h)$.

(ii) $\mathbb{E}u_i(\theta_i, \sigma_i^* \mid \sigma^*_{-i}, \hat{\beta}^*) = \sum_{(\theta_{-i},y,h) \in \Theta_{-i} \times \overline{Y} \times M} u_i(\theta_i, \theta_{-i}, y, h_i) \cdot \mathcal{M}(\theta_i, \theta_{-i})(y,h) \cdot \beta_i(\theta_i)(\theta_{-i})$.

(iii) If types are independent, then

$$\mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma^*_{-i}, \hat{\beta}^*) = \sum_{(\theta_{-i},y,h) \in \Theta_{-i} \times \overline{Y} \times M} u_i(\theta_i, \theta_{-i}, y, h_i) \cdot \mathcal{M}(\theta_i', \theta_{-i})(y,h) \cdot \operatorname{marg}_{\Theta_{-i}} \mu(\theta_{-i}).$$

Proof. Write $\hat{\delta}^*$ for the hierarchy mappings associated to $\hat{\beta}^*$. To show (i), first observe that

$$\begin{aligned}\mathbb{E}\pi(\sigma^*, \hat{\beta}^*) &= \sum_{(\theta,y,h) \in \Theta \times \overline{Y} \times M} \pi(\theta, y, \hat{\delta}^*([\theta, \theta, y, h])) \cdot \mathcal{M}(\theta)(y,h) \cdot \mu(\theta) \\ &= \sum_{(\theta,y,h) \in \Theta \times \overline{Y} \times M} \pi(\theta, y, \hat{\delta}^*([\theta, \theta, y, h])) \cdot \phi(\theta, y, h).\end{aligned}$$

Thus, it is sufficient to show that $\phi(\theta, y, h) > 0$ implies $\hat{\delta}^*([\theta, \theta, y, h]) = h$. This follows from the fact that $\mathbb{P}([\theta, \theta, y, h] \mid \sigma^*) = \phi(\theta, y, h) > 0$, so $[\theta_i, \theta_i, y_i, h_i] \in \hat{\mathcal{P}}_i^+$ for each $i \in I$. Thus, Lemma 4.1 implies $\hat{\delta}^*([\theta, \theta, y, h]) = h$.

Notice, for each $\theta_i, \theta_i' \in \Theta_i$,

$$\mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma^*_{-i}, \hat{\beta}^*) = \sum_{\substack{(\theta_{-i},y,h) \\ \in \Theta_{-i} \times \overline{Y} \times M}} u_i(\theta_i, \theta_{-i}, y, \hat{\delta}_i^*([\theta_i, \theta_i', y_i, h_i])) \cdot \mathcal{M}(\theta_i', \theta_{-i})(y,h) \cdot \beta_i(\theta_i)(\theta_{-i}). \tag{18}$$

We now show (ii). Notice that $\beta_i(\theta_i)$ has full support. Thus, to show (ii), it suffices to show that $\mathcal{M}(\theta_i, \theta_{-i})(y,h) > 0$ implies $\hat{\delta}_i^*([\theta_i, \theta_i, y_i, h_i]) = h_i$. But notice, if $\mathcal{M}(\theta_i, \theta_{-i})(y,h) > 0$ then $[\theta_i, \theta_i, y_i, h_i] \in \mathcal{P}_i^+$ which implies $\hat{\delta}_i^*([\theta_i, \theta_i, y_i, h_i]) = h_i$. (See Lemma 4.1.)

We now show (iii). Notice, by Equation (18) it suffices to show that $\mathcal{M}(\theta_i', \theta_{-i})(y,h) > 0$ implies $\hat{\delta}_i^*([\theta_i, \theta_i', y_i, h_i]) = h_i$. But notice, if $\mathcal{M}(\theta_i', \theta_{-i})(y,h)$ then $[\theta_i', \theta_i', y_i, h_i] \in \mathcal{P}_i^+$. Since types are independent, this implies $\hat{\delta}_i^*([\theta_i, \theta_i', y_i, h_i]) = h_i$. (See Lemma 4.1.) Moreover, independence of the types also implies $\beta_i(\theta_i)(\theta_{-i}) = \operatorname{marg}_{\Theta_{-i}} \mu(\theta_{-i})$. □

Appendix D. Proofs of Section 5

Fix a Bayesian game (Γ, μ). Let $\hat{\beta}$ be a profile of terminal beliefs of (Γ, μ), and write $\hat{\delta}$ for its associated terminal hierarchies of beliefs. In addition, write $\hat{\mathcal{P}}[\theta] := \{\hat{\rho} \in \hat{\mathcal{P}} : \operatorname{proj}_\Theta(\hat{\rho}) = \theta\}$.

Lemma D.1.

(i) For each $\theta \in \Theta$, $\{\hat{\mathcal{P}}[\theta, y, h] : (y,h) \in \overline{Y} \times M\}$ is a partition of $\hat{\mathcal{P}}[\theta]$.

(ii) For each $(\theta_i, y_i, h_i) \in \Theta_i \times \overline{Y}_i \times M_i$, $\{\hat{\mathcal{P}}[\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i}] : (\theta_{-i}, y_{-i}, h_{-i}) \in \Theta_{-i} \times \overline{Y}_{-i} \times M_{-i}\}$ is a partition of $\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}$.

(iii) For each $(\theta_i, \theta_{-i}, y_i, h_i, h_{-i}) \in \Theta \times \overline{Y}_i \times M$, $\{\hat{\mathcal{P}}[\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i}] : y_{-i} \in \overline{Y}_{-i}\}$ is a partition of $\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]$.

Proof. First we show (i). Fix $\theta \in \Theta$. Notice that $\hat{\mathcal{P}}[\theta] = \bigcup_{(\theta,y,h) \in \Theta \times Y \times M} \hat{\mathcal{P}}[\theta, y, h]$. In addition, note that $(\theta, y, h) \neq (\theta, y', h')$ implies $\hat{\mathcal{P}}[\theta, y, h] \cap \hat{\mathcal{P}}[\theta, y', h'] = \emptyset$.

Parts (ii) and (iii) follow from analogous arguments. □

Write $\mathcal{S}(y, h \mid \theta, \hat{\beta}) := \cup_{\hat{\rho} \in \hat{\mathcal{P}}[\theta,y,h]} \mathcal{S}(\hat{\rho})$ for the strategy profiles that: (i) induce the material outcome y, and (ii) induce the hierarchy profile h, given that the type profile is θ, and terminal beliefs are $\hat{\beta}$.

Lemma D.2.

(i) For each $\theta \in \Theta$, $\{\mathcal{S}(y, h \mid \theta, \hat{\beta}) : (y,h) \in Y \times M\}$ is a partition of $S_0 \times S$.

(ii) For each $\theta \in \Theta$, $\{\mathcal{S}(\hat{\rho}) : \hat{\rho} \in \hat{\mathcal{P}}[\theta]\}$ is a partition of $S_0 \times S$.

(iii) For each $(\theta, y, h) \in \Theta \times \overline{Y} \times M$, $\{\mathcal{S}(\hat{\rho}) : \hat{\rho} \in \hat{\mathcal{P}}[\theta, y, h]\}$ is a partition of $\mathcal{S}(y, h \mid \theta, \hat{\beta})$.

Proof. To show (i), fix some $\theta \in \Theta$. Notice $S_0 \times S = \bigcup_{(y,h) \in \overline{Y} \times M} \mathcal{S}(y, h \mid \theta, \hat{\beta})$. In addition, each strategy profile (s_0, s) leads to a unique pair $(y,h) \in \overline{Y} \times M$, where $y = \operatorname{proj}_Y(\tau(\theta, s_0, s))$ and $h = \hat{\delta}(\tau(\theta, s_0, s))$. Thus, if $(y,h), (y',h') \in Y \times M$ with $(y,h) \neq (y',h')$, then $\mathcal{S}(y, h \mid \theta, \hat{\beta}) \cap \mathcal{S}(y', h' \mid \theta, \hat{\beta}) = \emptyset$.

Parts (ii) and (iii) follow from analogous arguments. □

Lemma D.3. *Fix a Bayesian equilibrium (σ, β) of a Bayesian game (Γ, μ). Let $\mathcal{M}$ be the induced extended direct mechanism. Then, for each $\theta \in \Theta$, $\mathcal{M}(\theta)(\cdot)$ is a well defined probability measure with support in $\overline{Y} \times M := \operatorname{proj}_Y(\hat{\mathcal{P}}) \times \hat{\delta}(\hat{\mathcal{P}})$.*

Proof. Fix $\theta \in \Theta$. It suffices to show that $\mathcal{M}(\theta)(\overline{Y} \times M) = 1$. Notice that

$$\sum_{(y,h) \in \overline{Y} \times M} \mathcal{M}(\theta)(y,h) = \sum_{(y,h) \in \overline{Y} \times M} \mathbb{P}[\hat{\mathcal{P}}[\theta, y, h] \mid \theta, \sigma] = \mathbb{P}[\hat{\mathcal{P}}[\theta] \mid \theta, \sigma] = 1,$$

where the second equality follows from the fact that $\{\hat{\mathcal{P}}[\theta, y, h] : (y,h) \in Y \times H\}$ is a partition of $\hat{\mathcal{P}}[\theta]$ (Lemma D.1) and the third from the fact that $\operatorname{Supp}(\mathbb{P}[\cdot \mid \theta, \sigma]) \subseteq \hat{\mathcal{P}}[\theta]$. □

Recall from the main text that $\hat{\delta}$ (resp. $\hat{\delta}^*$) are the terminal hierarchies of beliefs induced by $\hat{\beta}$ (resp. $\hat{\beta}^*$). Also recall from Appendix C that $\tau_i(\theta_i, s_0, s) \in \hat{\mathcal{P}}_i$ specifies the terminal information set of agent i, given that i is of type θ_i and the agents play (s_0, s).

Lemma D.4. *Fix a Bayesian equilibrium of (Γ, μ), viz. $(\sigma, \hat{\beta})$. Let $\mathcal{M}$ be the extended direct mechanism induced by $(\sigma, \hat{\beta})$ and assume that it satisfies BLV. Let $\hat{\delta}^*$ be the hierarchy mappings induced by honest beliefs β^* in $(\Gamma(\mathcal{M}), \mu)$. For each $\theta_i \in \Theta_i$ and each $(s_0, s) \in \mathcal{S}(y, \widetilde{h} \mid \theta_i', \theta_{-i}, \hat{\beta})$ with $p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) > 0$, $\hat{\delta}_i(\tau_i(\theta_i, s_0, s)) = \hat{\delta}_i^*([\theta_i, \theta_i', y_i, \widetilde{h}_i])$.*

Proof. Fix $\theta_i \in \Theta_i$. Write $h_i = \hat{\delta}_i(\tau_i(\theta_i, s_0, s))$ and $h_i^* = \hat{\delta}_i^*([\theta_i, \theta_i', y_i, \widetilde{h}_i])$. To show that $h_i = h_i^*$ it suffices to show the following:

(i) There is $c > 0$, so that, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$,

$$h_i^\infty(\theta_{-i}, h_{-i}) = c \, \tfrac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \, \widetilde{h}_i^\infty(\theta_{-i}, h_{-i}).$$

(ii) There is $c^* > 0$, so that, for each $(\theta_{-i}, h_{-i}) \in \Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$,

$$h_i^{*\infty}(\theta_{-i}, h_{-i}) = c^* \, \tfrac{\mu(\theta_i, \theta_{-i})}{\mu(\theta_i', \theta_{-i})} \, \widetilde{h}_i^\infty(\theta_{-i}, h_{-i}).$$

To see why, note that $\operatorname{Supp} h_i^{\infty}$ and $\operatorname{Supp} h_i^{*\infty}$ are finite subsets of $\Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$. (See Lemma C.4.) Thus, (i) and (ii) imply h_i^{∞} is a multiple of $h_i^{*\infty}$. Since their sum over $\Theta_{-i} \times \hat{\delta}_{-i}(\hat{\mathcal{P}}_{-i})$ adds to 1, this implies $h_i^{\infty} = h_i^{*\infty}$, so $h_i = h_i^*$.

To prove (i), recall that $p_0(s_0) \cdot \sigma_{-i}(\theta_{-i})(s_{-i}) > 0$ and $\hat{\delta}_i(\tau_i(\theta_i, s_0, s)) = h_i$. Moreover, using the fact that $(s_0, s) \in \mathcal{S}(y, \widetilde{h} \mid \theta_i', \theta_{-i}, \hat{\beta})$ it follows that $\hat{\delta}_i(\tau_i(\theta_i', s_0, s)) = \widetilde{h}_i$. Therefore, Lemma C.5 establishes (i).

To prove (ii), first we show $\hat{\delta}_i^*([\theta_i', \theta_i', y_i, \widetilde{h}_i]) = \widetilde{h}_i$. Note that

$$\mathcal{M}(\theta_i', \theta_{-i})(y, \widetilde{h}) = \sum_{(s_0', s') \in \mathcal{S}(y, \widetilde{h} \mid (\theta_i', \theta_{-i}), \hat{\beta})} p_0(s_0') \cdot \sigma(\theta_i', \theta_{-i})(s') \geq \sigma(\theta_i', \theta_{-i})(s) \cdot p_0(s_0) > 0, \tag{19}$$

where the first inequality follows from the fact that $(s_0, s) \in \mathcal{S}(y, \widetilde{h} \mid \theta_i', \theta_{-i}, \hat{\beta})$. Thus,

$$\operatorname{marg}_{\Theta_i \times Y_i \times M_i} \phi(\theta_i', y_i, \widetilde{h}_i) \geq \phi(\theta_i', \theta_{-i}, y, \widetilde{h}) = \mu(\theta_i', \theta_{-i}) \cdot \mathcal{M}(\theta_i', \theta_{-i}, y, \widetilde{h}) > 0. \tag{20}$$

Write $\Gamma(\mathcal{M})$ for the canonical extensive form where (S_0^*, p_0^*) are chance strategies, S^* the agents' strategy profiles, and $\hat{\mathcal{P}}_i^*$ for the terminal information sets. By Equation (20), $[\theta_i', \theta_i', y_i, \widetilde{h}_i] \in \hat{\mathcal{P}}_i^{*+}$. Thus, since $\mathcal{M}$ satisfies BLV and $\hat{\beta}^*$ is honest, $\hat{\delta}_i^*([\theta_i', \theta_i', y_i, \widetilde{h}_i]) = \widetilde{h}_i$. (See Lemma 4.1.) Note that Equation (19) implies that chance selects $(y, \widetilde{h})$ with positive probability given the profile (θ_i', θ_{-i}), i.e. there is some $s_0^* \in S_0^*$ so that $s_0^*(\theta_i', \theta_{-i}) = (y, \widetilde{h})$ and $p_0^*(s_0^*) > 0$. Fix such s_0^*. Write $\tau_i^* : \Theta_i \times S_0^* \times S^* \to \hat{\mathcal{P}}_i^*$ for the terminal belief mapping of i and notice that $\tau_i^*(\theta_i', s_0^*, (\theta_i', \theta_{-i})) = [\theta_i', \theta_i', y_i, h_i]$ and $\tau_i^*(\theta_i, s_0^*, (\theta_i', \theta_{-i})) = [\theta_i, \theta_i', y_i, \widetilde{h}_i]$. In summary:

(a) $\hat{\delta}_i^*(\tau_i^*(\theta_i', s_0^*, (\theta_i', \theta_{-i}))) = \hat{\delta}_i([\theta_i', \theta_i', y_i, \widetilde{h}_i]) = \widetilde{h}_i$,
(b) $\hat{\delta}_i^*(\tau_i^*(\theta_i, s_0^*, (\theta_i', \theta_{-i}))) = \hat{\delta}_i^*([\theta_i, \theta_i', y_i, \widetilde{h}_i]) = h_i^*$, and
(c) $p_0^*(s_0^*) \cdot \sigma_{-i}^*(\theta_{-i})(\theta_{-i}) > 0$.

Thus, (ii) follows from Lemma C.5. □

Proof of Theorem 5.1. Let $\mathcal{M}$ be the extended direct mechanism induced by $(\sigma, \hat{\beta})$. Write $\hat{\delta}$ for the terminal hierarchies of beliefs induced by $\hat{\beta}$. The proof of Theorem 5.1 is divided into three steps. The first step shows that $\mathcal{M}$ satisfies BLV. The second step shows that $\mathcal{M}$ satisfies BIC. The last step shows conditions (i), (ii), and (iii).

Step 1: $\mathcal{M}$ satisfies BLV. We show that for each $(\theta_i, y_i, h_i) \in \Theta_i \times Y_i \times M_i$,

$$h_i^{\infty}(\theta_{-i}, h_{-i}) \cdot \operatorname{marg}_{\Theta_i \times Y_i \times M_i} \phi(\theta_i, y_i, h_i) = \operatorname{marg}_{\Theta \times Y_i \times M} \phi(\theta_i, \theta_{-i}, y_i, h_i, h_{-i}). \tag{21}$$

First, we show that Equation (21) is equivalent to

$$\begin{aligned} h_i^{\infty}(\theta_{-i}, h_{-i}) & \sum_{s_i \in S_i} \mathbb{P}\left(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i} \mid s_i, \sigma_{-i}\right) \cdot \sigma_i(\theta_i)(s_i) \\ & = \sum_{s_i \in S_i} \mathbb{P}\left(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}] \mid s_i, \sigma_{-i}\right) \cdot \sigma_i(\theta_i)(s_i). \end{aligned} \tag{22}$$

To show the equivalence, notice that, for each $(\theta, y, h) \in \Theta \times \overline{Y} \times M$,

$$\begin{aligned} \phi(\theta, y, h) &= \mu(\theta) \cdot \mathcal{M}(\theta)(y, h) \\ &= \sum_{(s_0, s) \in \mathcal{S}(y, h \mid \theta, \hat{\beta})} \mu(\theta) \cdot p_0(s_0) \cdot \sigma(\theta)(s) \\ &= \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta, y, h]} \sum_{(s_0, s) \in \mathcal{S}(\hat{\rho})} \mu(\theta) \cdot p_0(s_0) \cdot \sigma(\theta)(s) \\ &= \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta, y, h]} \sum_{s_i \in S_i} \mathbb{P}(\hat{\rho} \mid s_i, \sigma_{-i}) \cdot \sigma_i(\theta_i)(s_i) \\ &= \sum_{s_i \in S_i} \mathbb{P}\left(\hat{\mathcal{P}}[\theta, y, h] \mid s_i, \sigma_{-i}\right) \cdot \sigma_i(\theta_i)(s_i), \end{aligned} \tag{23}$$

where the second equality follows from the definition of $\mathcal{M}$, the third equality follows from the fact that $\{\mathcal{S}(\hat{\rho}) : \hat{\rho} \in \hat{\mathcal{P}}[\theta, y, h]\}$ is a partition of $\mathcal{S}(y, h \mid \theta, \hat{\beta})$ (Lemma D.2), and the fourth equality follows from the definition of $\mathbb{P}(\cdot)$. Thus,

$$\begin{aligned} \operatorname{marg}_{\Theta_i \times \overline{Y}_i \times M_i} \phi(\theta_i, y_i, h_i) &= \sum_{\substack{(\theta_{-i}, y_{-i}, h_{-i}) \\ \in \Theta_{-i} \times \overline{Y}_{-i} \times M_{-i}}} \phi(\theta_i, \theta_{-i}, y, y_{-i}, h_i, h_{-i}) \\ &= \sum_{\substack{(\theta_{-i}, y_{-i}, h_{-i}) \\ \in \Theta_{-i} \times \overline{Y}_{-i} \times M_{-i}}} \sum_{s_i \in S_i} \mathbb{P}\left(\hat{\mathcal{P}}[\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i}] \mid s_i, \sigma_{-i}\right) \cdot \sigma_i(\theta_i)(s_i) \end{aligned}$$

$$= \sum_{s_i \in S_i} \left(\mathbb{P}(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i} \mid s_i, \sigma_{-i}) \right) \cdot \sigma_i(\theta_i)(s_i), \tag{24}$$

where the second equality follows from Equation (23) and the last from the fact that the collection $\{\hat{\mathcal{P}}[\theta_i, \theta_{-i}, y_i, y_{-i}, h_i, h_{-i}] : (\theta_{-i}, y_{-i}, h_{-i}) \in \Theta_{-i} \times \overline{Y}_{-i} \times M_{-i}\}$ is a partition of the set $\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}$. (See Lemma D.1.) Similarly,

$$\begin{aligned} \operatorname{marg}_{\Theta \times \overline{Y}_i \times M} \phi(\theta, y_i, h) &= \sum_{y_{-i} \in \overline{Y}_{-i}} \phi(\theta, y_i, y_{-i}, h) \\ &= \sum_{y_{-i} \in \overline{Y}_{-i}} \sum_{s_i \in S_i} \mathbb{P}(\hat{\mathcal{P}}[\theta, y_i, y_{-i}, h] \mid s_i, \sigma_{-i}) \cdot \sigma_i(\theta_i)(s_i) \\ &= \sum_{s_i \in S_i} \left(\mathbb{P}(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}] \mid s_i, \sigma_{-i}) \right) \cdot \sigma_i(\theta_i)(s_i), \end{aligned} \tag{25}$$

where the second equality follows from Equation (23) and the last from the fact that $\{\hat{\mathcal{P}}[\theta, y_i, y_{-i}, h] : y_{-i} \in \overline{Y}_{-i}\}$ is a partition of the set $\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]$. (See Lemma D.1.) Thus, by Equations (24) - (25), Equation (21) is equivalent to Equation (22).

Now, to prove that Equation (22) holds, it suffices to show that, for each $s_i \in S_i$,

$$h_i^{\infty}(\theta_{-i}, h_{-i}) \cdot \mathbb{P}(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i} \mid s_i, \sigma_{-i}) = \mathbb{P}\left(\hat{\mathcal{P}}_i[\theta_i, y_i, h_i] \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}] \mid s_i, \sigma_{-i}\right). \tag{26}$$

To show Equation (26), fix $\hat{\rho}_i \in \hat{\mathcal{P}}_i[\theta_i, y_i, h_i]$ and notice that for each $s_i \in S_i$,

$$\begin{aligned} h_i^{\infty}(\theta_{-i}, h_{-i}) \cdot \mathbb{P}(\{\hat{\rho}_i\} \times \hat{\mathcal{P}}_{-i} \mid s_i, \sigma_{-i}) &= \hat{\beta}_i(\hat{\rho}_i)(\hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]) \cdot \mathbb{P}(\{\hat{\rho}_i\} \times \hat{\mathcal{P}}_{-i} \mid s_i, \sigma_{-i}) \\ &= \sum_{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]} \hat{\beta}_i(\hat{\rho}_i)(\hat{\rho}_{-i}) \cdot \mathbb{P}(\{\hat{\rho}_i\} \times \hat{\mathcal{P}}_{-i} \mid s_i, \sigma_{-i}) \\ &= \sum_{\hat{\rho}_{-i} \in \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}]} \mathbb{P}((\hat{\rho}_i, \hat{\rho}_{-i}) \mid s_i, \sigma_{-i}) \\ &= \mathbb{P}(\{\hat{\rho}_i\} \times \hat{\mathcal{P}}_{-i}[\theta_{-i}, h_{-i}] \mid s_i, \sigma_{-i}), \end{aligned}$$

where the first equality follows from Lemma C.4 and the third from consistency of $(\sigma, \hat{\beta})$. This establishes (26) and therefore $\mathcal{M}$ satisfies BLV.

Step 2: $\mathcal{M}$ satisfies BIC. Let $\hat{\beta}^*$ be honest terminal beliefs of the canonical Bayesian game. Since $(\sigma, \hat{\beta})$ is a Bayesian equilibrium, then for each $\theta_i, \theta_i' \in \Theta_i$,

$$\sum_{s_i \in S_i} \mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta}) \cdot \sigma_i(\theta_i)(s_i) \geq \sum_{s_i \in S_i} \mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta}) \cdot \sigma_i(\theta_i')(s_i).$$

Fix $\theta_i, \theta_i' \in \Theta_i$. We show that

$$\mathbb{E}u_i^*(\theta_i, \theta_i' \mid \sigma_{-i}^*, \hat{\beta}^*) = \sum_{s_i \in S_i} \mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta}) \cdot \sigma_i(\theta_i')(s_i). \tag{27}$$

From this, it follows that $\mathbb{E}u_i^*(\theta_i, \theta_i \mid \sigma_{-i}^*, \hat{\beta}^*) \geq \mathbb{E}u_i^*(\theta_i, \theta_i' \mid \sigma_{-i}^*, \hat{\beta}^*)$; so, $\mathcal{M}$ satisfies BIC.

Let $\hat{\delta}$ (resp $\hat{\delta}^*$) be the terminal hierarchies of beliefs induced by $\hat{\beta}$ (resp $\hat{\beta}^*$). The key is showing a relation between $\hat{\delta}$ and $\hat{\delta}^*$. Notice, by Lemma D.4, for each $(s_0, s) \in \hat{S}(y, h \mid \theta_i', \theta_{-i}, \hat{\beta})$ with $p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) > 0$,

$$\hat{\delta}_i(\tau_i(\theta_i, s_0, s)) = \hat{\delta}_i^*([\theta_i, \theta_i', y_i, h_i]). \tag{28}$$

Thus, on the one hand,

$$\begin{aligned} &\mathbb{E}u_i^*(\theta_i, \theta_i' \mid \sigma_{-i}^*, \hat{\beta}^*) \\ &= \sum_{(\theta_{-i}, y, h) \in \Theta_{-i} \times \overline{Y} \times M} u_i(\theta_i, \theta_{-i}, y, \hat{\delta}_i^*[\theta_i, \theta_i', y_i, h_i]) \cdot \mathcal{M}(\theta_i', \theta_{-i})(y, h) \cdot \beta_i(\theta_i)(\theta_{-i}) \\ &= \sum_{\substack{(\theta_{-i}, y, h) \in \Theta_{-i} \times \overline{Y} \times M \\ (s_0, s) \in \hat{S}(y, h \mid \theta_i', \theta_{-i}, \hat{\beta})}} u_i(\theta_i, \theta_{-i}, y, \hat{\delta}_i^*[\theta_i, \theta_i', y_i, h_i]) \cdot p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) \cdot \beta(\theta_i)(\theta_{-i}) \\ &= \sum_{\substack{(\theta_{-i}, y, h) \in \Theta_{-i} \times \overline{Y} \times M \\ (s_0, s) \in \hat{S}(y, h \mid \theta_i', \theta_{-i}, \hat{\beta})}} u_i(\theta_i, \theta_{-i}, \gamma(\zeta(s_0, s)), \hat{\delta}_i(\tau_i(\theta_i, s_0, s))) \cdot p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) \cdot \beta(\theta_i)(\theta_{-i}) \\ &= \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{(s_0, s) \in S_0 \times S} u_i(\theta_i, \theta_{-i}, \gamma(\zeta(s_0, s)), \hat{\delta}_i(\tau_i(\theta_i, s_0, s))) \cdot p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) \cdot \beta(\theta_i)(\theta_{-i}), \end{aligned}$$

where the second equality follows from the definition of $\mathcal{M}$, the third from Equation (28), and the last from the fact that $\{\hat{S}(y,h \mid \theta_i', \theta_{-i}, \hat{\beta}) : (y,h) \in \overline{Y} \times M\}$ is a partition of $S_0 \times S$ (Lemma D.2). On the other hand,

$$\begin{aligned}
&\sum_{s_i \in S_i} \mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta}) \cdot \sigma_i(\theta_i')(s_i) \\
&\quad = \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta_i, \theta_{-i}]} \sum_{s_i \in S_i} u_i(\text{proj}_\Theta(\rho), \text{proj}_Y(\hat{\rho}), \hat{\delta}_i(\hat{\rho})) \cdot \mathbb{P}[\hat{\rho} \mid \theta_i, s_i, \sigma_{-i}] \cdot \sigma_i(\theta_i')(s_i) \\
&\quad = \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta_i, \theta_{-i}]} \sum_{(s_0,s) \in S(\hat{\rho})} u_i(\text{proj}_\Theta(\rho), \text{proj}_Y(\hat{\rho}), \hat{\delta}_i(\hat{\rho})) \cdot p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) \cdot \beta(\theta_i)(\theta_{-i}) \\
&\quad = \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta_i, \theta_{-i}]} \sum_{(s_0,s) \in S(\hat{\rho})} u_i(\theta_i, \theta_{-i}, \gamma(\zeta(s_0,s)), \hat{\delta}_i(\tau_i(\theta_i, s_0, s))) \cdot p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) \cdot \beta(\theta_i)(\theta_{-i}) \\
&\quad = \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{(s_0,s) \in S_0 \times S} u_i(\theta_i, \theta_{-i}, \gamma(\zeta(s_0,s)), \hat{\delta}_i(\tau_i(\theta_i, s_0, s))) \cdot p_0(s_0) \cdot \sigma(\theta_i', \theta_{-i})(s) \cdot \beta(\theta_i)(\theta_{-i}),
\end{aligned}$$

where the last equation from the fact that $\{S[\hat{\rho}] : \hat{\rho} \in \hat{\mathcal{P}}[\theta_i, \theta_{-i}]\}$ is a partition of $S_0 \times S$. (See Lemma D.2.) This shows Equation (27).

Step 3: conditions (i), (ii), and (iii). To show (i), recall that steps 1 and 2 state that $\mathcal{M}$ satisfies BLV and BIC. Thus, by Lemma C.6, each honest profile $(\sigma^*, \hat{\beta}^*)$ is a Bayesian equilibrium.

To show (ii), notice that

$$\mathbb{E}u_i^*(\theta_i, \sigma_i^* \mid \sigma_{-i}^*, \hat{\beta}^*) = \sum_{s_i \in S_i} \mathbb{E}u_i(\theta_i, s_i \mid \sigma_{-i}, \hat{\beta}) \cdot \sigma_i(\theta_i)(s_i) = \mathbb{E}u_i(\theta_i, \sigma_i \mid \sigma_{-i}, \hat{\beta}),$$

where the first equality follows from Equation (27).

To show (iii), notice that

$$\begin{aligned}
\mathbb{E}\pi^*(\sigma^*, \hat{\beta}^*) &= \sum_{(\theta,y,h) \in \Theta \times \overline{Y} \times M} \pi(\theta, y, h) \cdot \phi(\theta, y, h) \\
&= \sum_{(\theta,y,h) \in \Theta \times \overline{Y} \times M} \pi(\theta, y, h) \cdot \mu(\theta) \cdot \mathcal{M}(y, h) \\
&= \sum_{(\theta,y,h) \in \Theta \times \overline{Y} \times M} \pi(\theta, y, h) \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta,y,h]} \mathbb{P}(\hat{\rho} \mid \sigma) \\
&= \sum_{(\theta,y,h) \in \Theta \times \overline{Y} \times M} \sum_{\hat{\rho} \in \hat{\mathcal{P}}[\theta,y,h]} \pi(\text{proj}_\Theta(\hat{\rho}), \text{proj}_Y(\hat{\rho}), \hat{\delta}(\hat{\rho})) \cdot \mathbb{P}(\hat{\rho} \mid \sigma) \\
&= \sum_{\hat{\rho} \in \hat{\mathcal{P}}} \pi(\text{proj}_\Theta(\hat{\rho}), \text{proj}_Y(\hat{\rho}), \hat{\delta}(\hat{\rho})) \cdot \mathbb{P}(\hat{\rho} \mid \sigma) \\
&= \mathbb{E}\pi(\sigma, \hat{\beta}),
\end{aligned}$$

where the first equality follows from Lemma C.7, the second from definition of ϕ, the third from definition of $\mathcal{M}$, and the fifth from the fact that $\{\hat{\mathcal{P}}[\theta,y,h] : (\theta,y,h) \in \Theta \times \overline{Y} \times H\}$ is a partition of $\hat{\mathcal{P}}$. (See Lemma D.1.) □

Appendix E. Proofs of Section 6

Lemma E.1. *Fix an extended direct mechanism $\mathcal{M}$ that satisfies BLV and has outcome mappings $(Q_i, T_i, F_i)_{i \in I}$. Under each honest profile (σ^*, β^*), the expected utility of i with valuation θ_i and report θ_i' is $\mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma_{-i}^*, \beta_i^*) = Q_i(\theta_i') \cdot \theta_i - T_i(\theta_i') + F_i(\theta_i')$.*

Proof. Since types are independent,

$$\begin{aligned}
\mathbb{E}u_i(\theta_i, \theta_i' \mid \sigma_{-i}^*, \beta_i^*) &= \sum_{\theta_{-i} \in \Theta_{-i}} \sum_{(x,t) \in Y} \sum_{h \in M} \big(\theta_i \cdot x_i - t_i + f(h_i)\big) \cdot \mathcal{M}(\theta_i', \theta_{-i})((x,t), h) \cdot \text{marg}_{\Theta_{-i}} \mu(\theta_{-i}) \\
&= Q_i(\theta_i') \cdot \theta_i - T_i(\theta_i') + F_i(\theta_i'),
\end{aligned}$$

where the first equality follows from Lemma C.7 (iii). □

Proof of Proposition 6.1. Fix an extended direct mechanism $\mathcal{M}$ satisfying BLV and BIC. Write $\widetilde{T}_i(\theta_i) := T_i(\theta_i) - F_i(\theta_i)$.

Let $\widetilde{\mathcal{M}}$ be a mechanism with interim allocation rules $(Q_i(\cdot))_{i \in I}$, transfer rules $(\widetilde{T}_i(\cdot))_{i \in I}$ in an environment with no belief-dependent preferences (i.e., $\widetilde{F}_i(\cdot) = 0$).

Observe that $U_i(\theta_i,\theta_i') = \theta_i \cdot Q_i(\theta_i') - T_i(\theta_i') + F_i(\theta_i') = \theta_i \cdot Q_i(\theta_i') - \widetilde{T}_i(\theta_i')$. Hence, $\mathcal{M}$ is BIC if and only if the mechanism $\widetilde{\mathcal{M}}$ is BIC in a setting with no belief-dependent preferences. Thus, by Lemma 1 in Bergemann and Pesendorfer (2007), the allocation rules $(Q_i(\cdot))_{i\in I}$ are weakly increasing. In addition, there are some numbers $(\tilde{v}_i^1,\dots,\tilde{v}_i^K)\in\mathbb{R}^K$ satisfying:

(i) $\tilde{v}_i^k = v_i(\theta_i^k)$, whenever $Q_i(\theta_i^k)=0$ or $k=K$.

(ii) $v_i(\theta_i^k) \geq \tilde{v}_i^k \geq \theta_i^k - (\theta_i^{k+1}-\theta_i^k)\cdot \frac{1-\sum_{j=1}^k \overline{\mu}_i(\theta_i^j)}{\overline{\mu}_i(\theta_i^k)} \cdot \frac{Q_i(\theta_i^{k+1})}{Q_i(\theta_i^k)}$, whenever $Q_i(\theta_i^k)\neq 0$ and $k<K$.

(iii) $\sum_{k=1}^K \widetilde{T}_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k) = \sum_{k=1}^K Q_i(\theta_i^k)\cdot\tilde{v}_i^k\cdot\overline{\mu}_i(\theta_i^k) - U_i(\theta_i^1,\theta_i^1)$.

Moreover, (iii) implies

$$\mathrm{Rev}_i(\mathcal{M}) = \sum_{k=1}^K T_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k) = \sum_{k=1}^K Q_i(\theta_i^k)\cdot\tilde{v}_i^k\cdot\overline{\mu}_i(\theta_i^k) - U_i(\theta_i^1,\theta_i^1) + \sum_{k=1}^K F_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k),$$

as desired. □

Lemma E.2. *For each $\mathcal{M}\in$ IMP, $\mathrm{Rev}(\mathcal{M}) \leq \mathrm{VW}(\mathcal{M}) + \mathrm{PW}(\mathcal{M})$. Moreover, the inequality binds if the transfers of $\mathcal{M}$ are maximally compatible.*

Proof. Fix $\mathcal{M}\in$ IMP. Notice, Proposition 6.1 implies that

$$\mathrm{Rev}_i(\mathcal{M}) \leq \sum_{k=1}^K Q_i(\theta_i^k)\cdot v_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k) + \sum_{k=1}^K F_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k).$$

This follows from the fact that $\tilde{v}_i^k \leq v_i(\theta_i^k)$ and that, in any IR mechanism, $U_i(\theta_i^1,\theta_i^1)\geq 0$. Therefore,

$$\sum_{i\in I}\mathrm{Rev}_i(\mathcal{M}) \leq \sum_{i\in I}\sum_{\theta_i\in\Theta_i} Q_i(\theta_i)\cdot v_i(\theta_i)\cdot\overline{\mu}_i(\theta_i) + \sum_{i\in I}\sum_{\theta_i\in\Theta_i} F_i(\theta_i)\cdot\overline{\mu}_i(\theta_i) = \mathrm{VW}(\mathcal{M}) + \mathrm{PW}(\mathcal{M}).$$

Now assume that $\mathcal{M}$ has maximally compatible transfers. Observe that

$$\begin{aligned}\sum_{k=2}^K\sum_{\ell=1}^{k-1}(\theta_i^{\ell+1}-\theta_i^\ell)\cdot Q_i(\theta_i^\ell)\cdot\overline{\mu}_i(\theta_i^k) &= \sum_{\ell=1}^{K-1}\sum_{k=\ell+1}^K(\theta_i^{\ell+1}-\theta_i^\ell)\cdot Q_i(\theta_i^\ell)\cdot\overline{\mu}_i(\theta_i^k)\\ &= \sum_{\ell=1}^{K-1}\left(1-\sum_{k=1}^{\ell}\overline{\mu}_i(\theta_i^k)\right)(\theta_i^{\ell+1}-\theta_i^\ell)\cdot Q_i(\theta_i^\ell),\end{aligned}$$

where the first equality follows from changing the order of the sums. Hence, by definition of the virtual value $v_i(\theta_i^k)$,

$$\sum_{k=1}^K\theta_i^k\cdot Q_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k) + \sum_{k=2}^K\sum_{\ell=1}^{k-1}(\theta_i^{\ell+1}-\theta_i^\ell)\cdot Q_i(\theta_i^\ell)\cdot\overline{\mu}_i(\theta_i^k) = \sum_{k=1}^K v_i(\theta_i^k)\cdot Q_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k). \tag{29}$$

Therefore, for each $i\in I$,

$$\begin{aligned}\sum_{k=1}^K T_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k) &= \sum_{k=1}^K\theta_i^k\cdot Q_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k) + \sum_{k=2}^K\sum_{\ell=1}^{k-1}(\theta_i^{\ell+1}-\theta_i^\ell)\cdot Q_i(\theta_i^\ell)\cdot\overline{\mu}_i(\theta_i^k) + \sum_{k=1}^K F_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k)\\ &= \sum_{k=1}^K v_i(\theta_i^k)\cdot Q_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k) + \sum_{k=1}^K F_i(\theta_i^k)\cdot\overline{\mu}_i(\theta_i^k),\end{aligned}$$

where the first equality follows from the fact that transfers are maximally compatible and the second from Equation (29). Therefore, $\mathrm{Rev}(\mathcal{M}) = \mathrm{VW}(\mathcal{M}) + \mathrm{PW}(\mathcal{M})$, as desired. □

Proof of Corollary 6.1. Let $\mathcal{M}\in$ IMP be a mechanism with maximally-compatible transfers that satisfies $\mathrm{VW}(\mathcal{M}) + \mathrm{PW}(\mathcal{M}) = \sup_{\mathcal{M}'\in\mathrm{IMP}}(\mathrm{VW}(\mathcal{M}') + \mathrm{PW}(\mathcal{M}'))$. Observe that, by Lemma E.2, $\mathrm{Rev}(\mathcal{M}) = \mathrm{VW}(\mathcal{M}) + \mathrm{PW}(\mathcal{M})$. Moreover, for each $\mathcal{M}'\in$ IMP, $\mathrm{VW}(\mathcal{M}') + \mathrm{PW}(\mathcal{M}') \geq \mathrm{Rev}(\mathcal{M}')$. Therefore,

$$\mathrm{Rev}(\mathcal{M}) = \mathrm{VW}(\mathcal{M}) + \mathrm{PW}(\mathcal{M}) = \sup_{\mathcal{M}'\in\mathrm{IMP}}(\mathrm{VW}(\mathcal{M}') + \mathrm{PW}(\mathcal{M}')) \geq \sup_{\mathcal{M}'\in\mathrm{IMP}}\mathrm{Rev}(\mathcal{M}'),$$

as desired. □

Lemma E.3. *If $\mathcal{M}$ is direct mechanism satisfying BLV with weakly increasing allocation rules $(Q_i(\cdot))_{i\in I}$ and maximally compatible transfers $(T_i(\cdot))_{i\in I}$, then $\mathcal{M}$ is BIC and IR.*

Proof. First we show that $\mathcal{M}$ satisfies BIC. Write $\overline{T}_i(\theta_i) := T_i(\theta_i) - F_i(\theta_i)$. It suffices to show that for each $\theta_i, \theta_i' \in \Theta_i$, $\theta_i \cdot Q_i(\theta_i) - \overline{T}_i(\theta_i) \geq \theta_i \cdot Q_i(\theta_i') - \overline{T}_i(\theta_i')$, or equivalently, that $\theta_i(Q_i(\theta_i) - Q_i(\theta_i')) \geq \overline{T}_i(\theta_i') - \overline{T}_i(\theta_i)$. So, it suffices to show that for each $1 \leq k < \ell \leq K$,

$$\theta_i^{\ell}(Q_i(\theta_i^{\ell}) - Q_i(\theta_i^k)) \geq \overline{T}_i(\theta_i^{\ell}) - \overline{T}_i(\theta_i^k) \geq \theta_i^k(Q_i(\theta_i^{\ell}) - Q_i(\theta_i^k)).$$

Note that, by definition, for each $1 \leq j < K$, $\overline{T}_i(\theta_i^{j+1}) - \overline{T}_i(\theta_i^j) = \theta_i^{j+1}(Q_i(\theta_i^{j+1}) - Q_i(\theta_i^j))$, with $Q_i(\theta_i^{k+1}) - Q_i(\theta_i^k) \geq 0$. Thus, for each $1 \leq k < \ell \leq K$,

$$\overline{T}_i(\theta_i^{\ell}) - \overline{T}_i(\theta_i^k) = \sum_{j=k}^{\ell-1} \theta_i^j(Q_i(\theta_i^{j+1}) - Q_i(\theta_i^j)) \geq \sum_{j=k}^{\ell-1} \theta_i^k(Q_i(\theta_i^{j+1}) - Q_i(\theta_i^j)) = \theta_i^k(Q_i(\theta_i^k) - Q_i(\theta_i^j)).$$

Using an analogous argument $\overline{T}_i(\theta_i^{\ell}) - \overline{T}_i(\theta_i^k) \leq \theta_i^{\ell}(Q_i(\theta_i^k) - Q_i(\theta_i^j))$.

To show that $\mathcal{M}$ satisfies IR, we show that $U_i(\theta_i^k, \theta_i^k) \geq 0$ for each $\theta_i^k \in \Theta_i$. Note, since transfers are maximally compatible, $Q_i(\theta_i^1) \cdot \theta_i^1 - T_i(\theta_i^1) + F_i(\theta_i^1) = 0$. Thus,

$$U_i(\theta_i^k, \theta_i^k) \geq U_i(\theta_i^k, \theta_i^1) = Q_i(\theta_i^1) \cdot \theta_i^k - T_i(\theta_i^1) + F_i(\theta_i^1) \geq Q_i(\theta_i^1) \cdot \theta_i^1 - T_i(\theta_i^1) + F_i(\theta_i^1) = 0,$$

where the first inequality follows from BIC and the second from the fact that $\theta_i^k \geq \theta_i^1$. □

E.1. Image concerns

Fix agents $i, j \in I$ with $i \neq j$ and write $\mathbb{E}_j[\theta_i \mid h_j] := \sum_{\theta_i \in \Theta_i} \theta_i \cdot \text{marg}_{\Theta_i} h_j^{\infty}(\theta_i)$. Write $\mathbb{E}_i[\mathbb{E}_j[\theta_i] \mid h_j] := \int_{H_j} \mathbb{E}_j[\theta_i \mid h_j]\, d\text{marg}_{H_j} h_i^{\infty}$. Observe, since $\mathbb{E}_j[\theta_i \mid \cdot\,]$ is measurable and bounded, the mapping $\mathbb{E}_i[\mathbb{E}_j[\theta_i] \mid \cdot\,] : H_i \to \mathbb{R}$ is measurable. (See Theorem 15.13 in Aliprantis and Border (2006).) Say i has **expectation-based image concerns** if i's belief-based payoff is given by $f_i(h_i) = a \sum_{j \in I \setminus \{i\}} \mathbb{E}_i[\mathbb{E}_j[\theta_i] \mid h_i]$ for some $a > 0$.

Fix $E_j \subseteq \Theta_j$. Set $\mathcal{B}_j^1(E_j) := H_j$. For $i \neq j$ set $\mathcal{B}_i^1(E_j) := \{h_i \in H_i : \text{marg}_{\Theta_j} h_i^{\infty}(E_j) = 1\}$. Fix $k \in \mathbb{N}$ and assume that $(\mathcal{B}_i^k(E_j))_{i \in I}$ are measurable sets. Then, inductively define the sets $\mathcal{B}_i^{k+1}(E_j) := \{h_i \in \mathcal{B}_i^k(E_j) : \text{marg}_{H_{-i}} h_i^{\infty}(\mathcal{B}_{-i}^k(E_j)) = 1\}$. Notice that each set $\mathcal{B}_i^k(E_j)$ is measurable. (See Corollary 15.6 in Aliprantis and Border (2006).) Write $\mathcal{B}_i(E_j) := \cap_{k=1}^{\infty} \mathcal{B}_i^k(E_j)$ and notice it is measurable. Notice, if $h_i \in \mathcal{B}_i(E_j)$, then $\text{marg}_{\Theta_j \times H_{-i}} h_i^{\infty}(E_j \times \mathcal{B}_{-i}(E_j)) = 1$ for each $i \neq j$. If $h_j \in \mathcal{B}_j(E_j)$, then $\text{marg}_{H_{-j}} h_j^{\infty}(\mathcal{B}_{-j}(E_j)) = 1$.[20] Call $\mathcal{B}(E_j) := \prod_{i \in I} \mathcal{B}_i(E_j)$ the set of hierarchy profiles where there is **common belief of** E_j.

Fix $b > 0$ and consider the set $E_i = \{\theta_i \in \Theta_i : \theta_i \geq b\}$. Say i has **sophisticated-type image concerns** if i belief-based payoff $f_i(h_i) = a \cdot \mathbb{1}[h_i \in \mathcal{B}_i(E_i)]$, for some $a > 0$. Notice, since $\mathcal{B}_i(E_i)$ is measurable, f_i is measurable. Now, consider the set $\widetilde{E}_i = \{\theta_i \in \Theta_i : \theta_i < b\}$. Say i has **unsophisticated-type image concerns** if i has belief-based payoff $f_i(h_i) = a \cdot \mathbb{1}[h_i \notin \mathcal{B}_i(\widetilde{E}_i)]$, for some $a > 0$. Notice, since $\mathcal{B}_i(\widetilde{E}_i)$ is measurable, f_i is measurable.

Fix an extended direct mechanism $\mathcal{M}$ with support $\overline{Y} \times M$ that satisfies BLV. Fix a set $E_j \subseteq \Theta_j$ and write $B_i^j := M_i \cap \mathcal{B}_i(E_j)$ and $B^j := \prod_{k \in I} B_k^j$. So, $B^j \subseteq M$ is the set hierarchy-message profiles where there is common belief of E_j. Say $\mathcal{M}$ **publicly reveals** the set E_j if $\text{marg}_M \mathcal{M}(\theta_j, \theta_{-j})(B^j) = 1$ for each $(\theta_j, \theta_{-j}) \in E_j \times \Theta_{-j}$. Say $\mathcal{M}$ **conceals** the set E_j if $B_i^j = \emptyset$ for each $i \in I$.

Lemma E.4. *Fix an extended direct mechanism $\mathcal{M}$ with support $\overline{Y} \times M$ and ex-ante probability measure ϕ. Fix $j \in I$ and $E_j \subseteq \Theta_j$. For each $i, \in I$, write $B_i^j := M_i \cap \mathcal{B}_i(E_j)$ and write $B^j = \prod_{i \in I} B_i^j$. If $\mathcal{M}$ satisfies BLV, then ϕ satisfies the following properties:*

(i) For each $\theta_i \in \Theta_i$, $\text{marg}_{\Theta_i} \phi(\theta_i) = \overline{\mu}_i(\theta_i)$.

(ii) For each $h_i \in M_i$, and each subset $A_{-i} \subseteq \Theta_{-i} \times M_{-i}$, $h_i^{\infty}(A_{-i}) \cdot \text{marg}_{M_i} \phi(h_i) = \text{marg}_{M_i \times (\Theta_{-i} \times M_{-i})} \phi(\{h_i\} \times A_{-i})$.

(iii) For each $i, j \in I$, $\text{marg}_{M_i} \phi(B_i^j) = \text{marg}_M \phi(B^j) = \text{marg}_{\Theta_i \times M} \phi(E_i \times B^j)$.

(iv) Fix $C_i \subseteq M_i$ for each $i \in I$ and write $C = \prod_{i \in I} C_i$. If for each $i \in I$, $\text{marg}_{M_i} \phi(C_i) = \text{marg}_{\Theta_j \times M} \phi(E_j \times C)$, then $C \subseteq B^j$.

(v) For each $i \in I$, $\sum_{h_i \in M_i} \mathbb{E}_i[\mathbb{E}_j[\theta_i] \mid h_i] \cdot \text{marg}_{M_i} \phi(h_i) = \sum_{\theta_i \in \Theta_i} \theta_i \cdot \overline{\mu}_i(\theta_i)$.

(vi) $\text{PW}(\mathcal{M}) = \sum_{i \in I} \sum_{h_i \in M_i} f_i(h_i) \cdot \text{marg}_{M_i} \phi(h_i)$.

Proof. Property (i). Fix $\theta \in \Theta$. By definition of ϕ, $\text{marg}_{\Theta} \phi(\theta) = \mathcal{M}(\theta)(\overline{Y} \times M) \cdot \mu(\theta) = \mu(\theta)$. Hence, for each $\theta_i \in \Theta_i$, $\text{marg}_{\Theta_i} \phi(\theta_i) = \overline{\mu}_i(\theta_i)$.

Property (ii). Notice that

$$\text{marg}_{M_i \times \Theta_{-i} \times M_{-i}} \phi(\{h_i\} \times A_{-i}) = \sum_{(\theta_{-i}, h_{-i}) \in A_{-i}} \sum_{(\theta_i, y_i) \in \Theta_i \times Y_i} \text{marg}_{\Theta \times Y_i \times M} \phi(\theta_i, \theta_{-i}, y_i, h_i, h_{-i})$$

[20] Recall that $h_j^{\infty} \in \Delta(\Theta_{-j} \times H_{-j})$. So, j's the domain of uncertainty does not include his type Θ_j.

$$= \sum_{(\theta_{-i},h_{-i})\in A_{-i}} \sum_{(\theta_i,y_i)\in\Theta_i\times Y_i} h_i^\infty(\theta_{-i},h_{-i})\cdot \operatorname{marg}_{\Theta_i\times Y_i\times M_i}\phi(\theta_i,y_i,h_i)$$
$$= \sum_{(\theta_{-i},h_{-i})\in A_{-i}} h_i^\infty(\theta_{-i}h_{-i}) \sum_{(\theta_i,y_i)\in\Theta_i\times Y_i} \operatorname{marg}_{\Theta_i\times Y_i\times M_i}\phi(\theta_i,y_i,h_i)$$
$$= \operatorname{marg}_{H_j} h_i^\infty(A_{-i})\cdot \operatorname{marg}_{M_i}\phi(h_i),$$

where the second equality follows from BLV.

Property (iii). We divide the proof in two steps:

Step 1. Fix $k\neq j$. First we show that $\operatorname{marg}_{M_k}\phi(B_k^j)=\operatorname{marg}_{\Theta_j\times M}\phi(E_j\times B^j)$. Notice that $h_k\in B_k^j$ implies $\operatorname{marg}_{\Theta_j\times H_{-i}} h_k^\infty(E_j\times B_{-k}^j)=1$. Hence,

$$\operatorname{marg}_{M_k}\phi(h_k)=\operatorname{marg}_{\Theta_j\times H_{-k}} h_k^\infty(E_j\times B_{-k}^j)\cdot\operatorname{marg}_{M_k}\phi(h_k)=\operatorname{marg}_{\Theta_j\times M_k\times M_{-k}}\phi(E_j\times\{h_k\}\times B_{-k}^j),$$

where the second equality follows from Property (ii). Adding the equation above over $h_k\in B_k^j$, implies $\operatorname{marg}_{M_k}\phi(B_k^j)=\operatorname{marg}_{\Theta_j\times M}\phi(E_j\times B^j)$.

Step 2. We show the result. Notice that, $h_j\in B_j^j$ implies $h_j^\infty(\Theta_{-j}\times B_{-i}^j)=1$ Hence,

$$\operatorname{marg}_{M_j}\phi(h_j)=h_j^\infty(\Theta_{-j}\times B_{-j}^j)\cdot\operatorname{marg}_{M_j}\phi(h_j)=\operatorname{marg}_{M}\phi(\{h_j\}\times B_{-j}^j),$$

where the second equality follows from Property (ii). Adding the equation above over $h_j\in B_j^j$, implies $\operatorname{marg}_{M_j}\phi(B_j^j)=\operatorname{marg}_{M}\phi(B^j)$. Thus,

$$\operatorname{marg}_{M_j}\phi(B_j^j)=\operatorname{marg}_{M}\phi(B^j)\leq\operatorname{marg}_{M_k}\phi(B_k^j)=\operatorname{marg}_{\Theta_j\times M}\phi(E_j\times B^j)\leq\operatorname{marg}_{M_i}\phi(B_j^j),$$

where the second equality follows from Step 1. Consequently, the inequalities above hold with equality, as desired.

Property (iv). We show the following:

(a) for each $i\neq j$ and each $h_i\in C_i$, $\operatorname{marg}_{\Theta_{j\times M_{-i}}} h_i^\infty(E_j\times C_{-i})=1$, and

(b) for each $h_j\in C_j$, $\operatorname{marg}_{H_{-j}} h_j^\infty(C_{-j})=1$.

Notice, (a) and (b) imply that $C_i\subseteq\mathcal{B}_i^1(E_j)$ for each $i\in I$. Moreover, by an inductive argument, if $C_i\subseteq\mathcal{B}_i^k(E_j)$ for each $i\in I$, then (a) and (b) imply $C_i\subseteq\mathcal{B}_i^{k+1}(E_j)$ for each $i\in I$. Consequently, (a) and (b) imply that each $C_i\subseteq\mathcal{B}_i(E_j)$.

We show (a). Fix $i\neq j$ and notice that

$$\sum_{h_i\in C_i}\operatorname{marg}_{M_i}\phi(h_i)=\sum_{h_i\in C_i}\operatorname{marg}_{\Theta_j\times M_i\times M_{-i}}\phi(E_j\times\{h_i\}\times C_{-i})=\sum_{h_i\in C_i}\operatorname{marg}_{M_i}\phi(h_i)\cdot h_i^\infty(E_j\times C_{-i}),$$

where the first equality follows from the fact that $\operatorname{marg}_{M_i}\phi(C_i)=\operatorname{marg}_{\Theta_j\times M}\phi(E_j\times C)$ and the last from Property (ii). Notice, each $h_i\in C_i$ satisfies $\phi(h_i)>0$ and $h_i^\infty(E_j\times C_{-j})\leq 1$. Thus, the equality above holds if and only if $h_i^\infty(E_j\times C_{-j})=1$ for each $h_i\in C_i$, as desired.

Part (b) follows from an analogous argument.

Property (v). Notice that,

$$\sum_{h_i\in M_i}\mathbb{E}_i[\mathbb{E}_j[\theta_i]\mid h_i]\cdot\operatorname{marg}_{M_i}\phi(h_i)=\sum_{(h_i,h_j,\theta_i)\in M_i\times M_j\times\Theta_i}\theta_i\cdot\operatorname{marg}_{\Theta_i}h_j^\infty(\theta_i)\cdot\operatorname{marg}_{M_j}h_i^\infty(h_j)\cdot\operatorname{marg}_{M_i}\phi(h_i)$$
$$=\sum_{(h_i,h_j,\theta_i)\in M_i\times M_j\times\Theta_i}\theta_i\cdot\operatorname{marg}_{\Theta_i}h_j^\infty(\theta_i)\cdot\operatorname{marg}_{M}\phi(h_i,h_j)$$
$$=\sum_{(h_j,\theta_i)\in M_j\times\Theta_i}\theta_i\cdot\operatorname{marg}_{\Theta_i}h_j^\infty(\theta_i)\cdot\operatorname{marg}_{M_j}\phi(h_j)$$
$$=\sum_{(h_j,\theta_i)\in M_j\times\Theta_i}\theta_i\cdot\operatorname{marg}_{\Theta_i\times M}\phi(\theta_i,h_j)$$
$$=\sum_{\theta_i\in\Theta_i}\theta_i\cdot\operatorname{marg}_{\Theta_i}\phi(\theta_i),$$

where the first equality follows from definition of $\mathbb{E}_i[\mathbb{E}_j[\theta_i]\mid\cdot\,]$, the second and fourth from Property (ii). Since $\operatorname{marg}_{\Theta_i}\phi(\theta_i)=\overline{\mu}_i(\theta_i)$ (Property (i)), the desired equality holds.

Property (vi). Observe that

$$\mathrm{PW}(\mathcal{M})=\sum_{i\in I}\sum_{\theta_i\in\Theta_i}F_i(\theta_i)\cdot\overline{\mu}_i(\theta_i)$$

$$= \sum_{i\in I}\sum_{h_i\in M_i} f_i(h_i)\cdot \sum_{\theta_i\in\Theta_i}\sum_{\theta_{-i}\in\Theta_{-i}} \operatorname{marg}_{M_i}\mathcal{M}(\theta_i,\theta_{-i})(h_i)\cdot\mu(\theta_i,\theta_{-i})$$
$$= \sum_{i\in I}\sum_{h_i\in M_i} f_i(h_i)\cdot \operatorname{marg}_{M_i}\phi(h_i),$$

where the second equality follows from definition of F_i and the last from definition of ϕ. □

Lemma E.5. *Assume that agents are ex-ante symmetric and that the agents' virtual valuations are strictly increasing. Fix $\mathcal{M}\in$ IMP. Then* $\mathrm{VW}(\mathcal{M}) = \sup_{\mathcal{M}'\in\mathrm{IMP}}\mathrm{VW}(\mathcal{M}')$ *if and only if $\mathcal{M}$ has virtual-value cutoff.*

Proof. This follows from Corollary 1 in Bergemann and Pesendorfer (2007). □

Lemma E.6. *Suppose that the agents have expectation-based image concerns with psychological intensity $a>0$. For each $\mathcal{M}\in$ IMP,* $\mathrm{PW}(\mathcal{M}) = a\cdot(n-1)\sum_{i\in I}\sum_{\theta_i\in\Theta_i}\theta_i\cdot\overline{\mu}_i(\theta_i)$.

Proof. Fix $\mathcal{M}\in$ IMP, let ϕ be its ex-ante distribution, and $Y\times M$ its support. Notice, by Lemma E.4 (vi),

$$\mathrm{PW}(\mathcal{M}) = \sum_{i\in I}\sum_{h_i\in M_i} f_i(h_i)\cdot\operatorname{marg}_{M_i}\phi(h_i) = \sum_{i,j\in I, j\neq i}\sum_{h_i\in M_i} a\cdot\mathbb{E}_i[\mathbb{E}_j[\theta_i]\mid h_i]\cdot\operatorname{marg}_{M_i}\phi(h_i).$$

Thus, the result follows from Lemma E.4 (v). □

Lemma E.7. *Suppose that agents have expectation-based image concerns, are ex-ante symmetric, and that the agents' virtual valuations are strictly increasing. The following hold:*

(i) There exists a revenue maximizing mechanism $\mathcal{M}\in$ IMP.
(ii) Any implementable revenue-maximizing mechanism has a virtual-value cutoff.
(iii) The information revealed to the agents does not change the auctioneer's expected revenue.

Proof. Let $\mathcal{M}$ be an extended direct mechanism that satisfies the following:
(a) $\mathcal{M}$ has a virtual-value cutoff.
(b) $\mathcal{M}$ has maximally compatible transfers.
(c) $\mathcal{M}$ satisfies BLV.

We first show (i). Notice that Lemma E.3 implies that $\mathcal{M}$ is BIC and IR. So, $\mathcal{M}\in$ IMP. Lemma E.5 implies that $\mathrm{VW}(\mathcal{M}) = \sup_{\mathcal{M}'\in\mathrm{IMP}}(\mathrm{VW}(\mathcal{M}'))$. By Lemma E.6, $\mathrm{PW}:\mathrm{IMP}\to\mathbb{R}$ is constant. Thus, $\mathrm{PW}(\mathcal{M}) = \sup_{\mathcal{M}'\in\mathrm{IMP}}(\mathrm{PW}(\mathcal{M}'))$. Therefore, $\mathcal{M}$ is revenue maximizing. (See Corollary 6.1).

Part (ii) follows from Lemma E.5 and Part (iii) follows from the fact that the psychological welfare $\mathrm{PW}:\mathrm{IMP}\to\mathbb{R}$ is constant. (See Lemma E.6). □

Lemma E.8. *Suppose that agents have sophisticated-type image concerns with psychological reward $a>0$ and threshold $b>0$, write $E_i = \{\theta_i\in\Theta_i : \theta_i\geq b\}$. If $\mathcal{M}\in$ IMP, then:*

(i) $\mathrm{PW}(\mathcal{M})\leq a\sum_{i\in I}\overline{\mu}_i(E_i)$.
(ii) $\mathrm{PW}(\mathcal{M}) = a\sum_{i\in I}\overline{\mu}_i(E_i)$ *if and only if $\mathcal{M}$ publicly reveals each set E_i.*

Proof. Fix $\mathcal{M}\in$ IMP. Write $B_i^j\subseteq M_i$ for the set of i's hierarchy messages where there is common belief of E_j and write $B^j = \prod_{i\in I}B_i^j$. First note that

$$\operatorname{marg}_{M_i}\phi(B_i^i) = \operatorname{marg}_{\Theta_i\times M}\phi(E_i\times B^i)\leq\operatorname{marg}_{\Theta_i}\phi(E_i) = \overline{\mu}_i(E_i), \tag{30}$$

where the first equality follows from Lemma E.4 (iii) and the last from Lemma E.4 (i). Thus,

$$\mathrm{PW}(\mathcal{M}) = \sum_{i\in I}\sum_{h_i\in M_i} f_i(h_i)\cdot\operatorname{marg}_{M_i}\phi(h_i) = \sum_{i\in I} a\cdot\operatorname{marg}_{M_i}\phi(B_i^i)\leq\sum_{i\in I} a\cdot\operatorname{marg}_{M_i}\overline{\mu}_i(E_i),$$

where the first equality follows from Lemma E.4 (vi), and the inequality follows from Equation (30). This shows Part (i).

Observe, $\mathrm{PW}(\mathcal{M}) = a\sum_{i\in I}\overline{\mu}_i(E_i)$ if and only if Equation (30) holds with equality for each $i\in I$. But notice, $\operatorname{marg}_{M_i}\phi(E_i\times B_i^i) = \overline{\mu}_i(E_i)$ holds if and only if

$$\sum_{(\theta_i,\theta_{-i})\in E_i\times\Theta_{-i}}\operatorname{marg}_{M_i}\mathcal{M}(\theta_i,\theta_{-i})(B^i)\cdot\mu(\theta_i,\theta_{-i}) = \sum_{(\theta_i,\theta_{-i})\in E_i\times\Theta_{-i}}\mu(\theta_i,\theta_{-i}),$$

or equivalently, if $(\theta_i,\theta_{-i})\in E_i\times\Theta_{-i}$ implies $\operatorname{marg}_M\mathcal{M}(\theta_i,\theta_{-i})(B^i) = 1$. Thus, $\mathrm{PW}(\mathcal{M}) = a\sum_{i\in I}\overline{\mu}_i(E_i)$ if and only if $\mathcal{M}$ publicly reveals each set E_i. Hence, part (ii) holds. □

Lemma E.9. *Suppose that agents have sophisticated-type image concerns, are ex-ante symmetric, and that the agents' virtual valuations are strictly increasing. The following hold:*

(i) There exists a revenue maximizing mechanism $\mathcal{M} \in \text{IMP}$.
(ii) Any implementable revenue-maximizing mechanism has a virtual-value cutoff.
(iii) Any implementable revenue-maximizing mechanism publicly reveals whether agent i *is above cutoff* b.

Proof. Write $E_i = \{\theta_i \in \Theta_i : \theta_i \geq b\}$. Let $\mathcal{M}$ be an extended direct mechanism such that:
(a) $\mathcal{M}$ has a virtual-value cutoff.
(b) $\mathcal{M}$ has maximally compatible transfers,
(c) $\mathcal{M}$ satisfies BLV, and
(d) $(y,h) \in \operatorname{Supp} \mathcal{M}(\theta)$ iff for each i there is common belief of each θ_i at message profile h.
We first show (i). Notice that Lemma E.3 implies that $\mathcal{M}$ is BIC and IR. So, $\mathcal{M} \in \text{IMP}$. Lemma E.5 implies that $\text{VW}(\mathcal{M}) = \sup_{\mathcal{M}' \in \text{IMP}} \text{VW}(\mathcal{M}')$. Notice that (d) implies that $\mathcal{M}$ publicly reveals each E_i. Hence, Lemma E.8 implies that $\text{PW}(\mathcal{M}) = \sup_{\mathcal{M}' \in \text{IMP}} \text{PW}(\mathcal{M})$. Therefore, $\mathcal{M}$ is revenue maximizing. (See Corollary 6.1.)

Finally, part (ii) follows from Lemma E.5 and part (iii) follows from Lemma E.8. □

Lemma E.10. *Assume each agent* i *has unsophisticated-type image concerns with belief-reward* $a > 0$ *and threshold* $b > 0$. *Fix* $\mathcal{M} \in \text{IMP}$. *The following hold:*

(i) $\text{PW}(\mathcal{M}) \leq a \cdot n$.
(ii) $\text{PW}(\mathcal{M}) = a \cdot n$ *if and only if* $\mathcal{M}$ *conceals each set* $E_i = \{\theta_i \in \Theta_i : \theta_i < b\}$.

Proof. Write $B_i^j := M_i \cap \mathcal{B}_i(E_j)$ and $B^j := \prod_{i \in I} B_i^j$. Notice, by Lemma E.4 (vi),

$$\text{PW}(\mathcal{M}) = \sum_{i \in I} \sum_{h_i \in M_i} f_i(h_i) \cdot \operatorname{marg}_{M_i} \phi(h_i) = \sum_{i \in I} a \cdot \operatorname{marg}_{M_i} \phi(M_i \setminus B_i^j) \leq n \cdot a.$$

Thus, (i) holds. Moreover, since each message is sent with positive probability, $\text{PW}(\mathcal{M}) = n \cdot a$ if and only if each set B_i^i is empty, i.e., if $\mathcal{M}$ conceals each set E_i. Thus, (ii) holds. □

Lemma E.11. *Suppose that agents have unsophisticated-type image concerns, are ex-ante symmetric, and that the agents' virtual valuations are strictly increasing. The following hold:*

(i) There exists a revenue maximizing mechanism $\mathcal{M} \in \text{IMP}$.
(ii) Any implementable revenue-maximizing mechanism has a virtual-value cutoff.
(iii) Any implementable revenue-maximizing mechanism conceals whether agent i *is below the cutoff* b.

Proof. Let $\mathcal{M}$ be an extended direct mechanism that satisfies the following:
(a) $\mathcal{M}$ has a virtual-value cutoff.
(b) $\mathcal{M}$ is symmetric in the following sense: If $\operatorname{marg}_Y \mathcal{M}(\theta)(W_i) > 0$ and $\operatorname{marg}_Y \mathcal{M}(\theta)(W_j) > 0$, then $\operatorname{marg}_Y \mathcal{M}(\theta)(W_i) = \operatorname{marg}_Y \mathcal{M}(\theta)(W_j)$.
(c) $\mathcal{M}$ has maximally compatible transfers.
(d) For each $\theta \in \Theta$, $(x_i, t_i) \in \operatorname{Supp}(\operatorname{marg}_{Y_i} m(\theta))$ if and only if $t_i = T_i(\theta_i)$.
(e) $\mathcal{M}$ satisfies BLV.
(f) For each $(\theta_i, y_i) \in \operatorname{Supp}(\operatorname{marg}_{\Theta_i \times Y_i} \phi)$ there is a unique hierarchy message $h_i \in H_i$ such that $\operatorname{marg}_{\Theta_i \times Y_i \times M_i} \phi(\theta_i, y_i, h_i) > 0$.
Conditions (a) and (b) describe how the object is allocated. Notice, (b) states that ties are solved by allocating the object with equal probability to the agents with the highest reports. Conditions (c) and (d) describe the transfers. Condition (d) states that i's transfer t_i is not informative about θ_{-i}. i.e. each report θ_i induces a transfer $T_i(\theta_i)$ independently of the reports θ_{-i}. Conditions (e) and (f) describe the hierarchy messages sent to the agents. Note, each positive-probability pair (θ_i, y_i) receives only one hierarchy-message $h_i \in M_i$. Hence, given the pair (θ_i, y_i), the messages sent by $\mathcal{M}$ do not provide any information to i. So, overall, Conditions (d)-(f) state that each agent i only observes whether he wins the object ($x = 1$) or not ($x = 0$). In particular i does not observe the reports of the other agents nor who (other than i) obtains the object.

For each $i, j \in I$, write $E_j = \{\theta_j \in \Theta_j : \theta_j < b\}$ and let $B_i^j := M_i \cap \mathcal{B}_i(E_j)$ be the set of i's hierarchy messages where there is common belief of E_j. First we show that $\mathcal{M}$ conceals the sets $(E_i)_{i \in I}$, i.e., we show that, B_i^j is empty for each $i, j \in I$. Steps 1 and 2 below derive properties of ϕ and Step 3 uses them to obtain a contradiction if $B_i^j \neq \emptyset$.

Step 1. We show that if $i \neq j$, then, $\operatorname{marg}_{\Theta_i \times M_i} \phi(E_i^c \times B_i^j) = 0$ and $\operatorname{marg}_{Y \times M_i} \phi(W_i^c \times B_i^j) = 0$. That is, if there is common belief that j's valuation is below benchmark, then i necessarily wins the object ($y \in W_i$) and i's valuation is below the benchmark ($\theta_i \in E_i$). We prove this by contrapositive: If $\phi(\theta, y, h_i, h_{-i}) > 0$ and $(\theta_i, y) \notin E_i \times W_i$, then $h_i \notin B_i^j$.

Fix $(\theta, y, h_i, h_{-i}) \in \operatorname{Supp} \phi$ such that $(\theta_i, y) \notin E_i \times W_i$. Notice, by Lemma E.4 (iii), $\operatorname{marg}_{\Theta_j \times M} \phi(E_j \times B^j) = \operatorname{marg}_M \phi(B^j)$ which implies $\operatorname{marg}_{\Theta_j \times M} \phi(E_j^c \times B^j) = 0$. Hence, to show that $h_i \notin B_i^j$, it suffices to show that $\operatorname{marg}_{\Theta_j \times M_i} \phi(E_j^c \times \{h_i\}) > 0$. There are two cases:
Case 1: $y \notin W_i$. In this case there is some $\theta_j \in \Theta_j$ such that $\theta_j \geq \max\{b, \theta_i\}$ that satisfies $\operatorname{marg}_{\Theta_j \times \Theta_i \times Y} \phi(\theta_j, \theta_i, y) > 0$. That is, if i

loses the auction, is possible that j's valuation is above b.

Case 2: $y \in W_i$ and $\theta_i \notin E_i$, i.e., $(\theta_i \geq b)$. In this case there is some $\theta_j \in \Theta_j$ with $\theta_i \geq \theta_j \geq b$ that satisfies $\text{marg}_{\Theta_j \times \Theta_i \times Y} \phi(\theta_j, \theta_i, y) > 0$. That is, if i wins the auction and i's valuation is above b, then is possible that j's valuation is also above b.

So, in each case there is some $\theta_j \in E_j^c$ such that $\text{marg}_{\Theta_j \times \Theta_i \times Y} \phi(\theta_j, \theta_i, y) > 0$. Thus, by Condition (f), $\text{marg}_{\Theta_j \times \Theta_i \times Y_i \times M_i} \phi(\theta_j, \theta_i, y_i, h_i) > 0$. Therefore, $\text{marg}_{\Theta_j \times M_i} \phi(E_j^c \times \{h_i\}) > 0$, as desired.

Step 2. We show that $B^i = B^j$ for each $i, j \in I$. So, whenever there is common belief of E_j, there is also common belief of E_i. Fix $i \neq j$. We show $B^j \subseteq B^i$. To do so, it suffices to show that for each $k \in I$, $\text{marg}_{M_k} \phi(B_k^j) = \text{marg}_{\Theta_i \times M} \phi(E_i \times B^j)$. (See by Lemma E.4 (iv).)

Notice, by Step 1, $\text{marg}_{\Theta_i \times M_i} \phi(E_i^c \times B_i^j) = 0$. Hence, $\text{marg}_{\Theta_i \times M} \phi(E_i^c \times B^j) = 0$, which implies $\text{marg}_M \phi(B^j) = \text{marg}_{\Theta_i \times M} \phi(E_i \times B^j)$. Moreover, for each $k \in I$,

$$\text{marg}_{M_k} \phi(B_k^j) = \text{marg}_M \phi(B^j) = \text{marg}_{\Theta_i \times M} \phi(E_i \times B^j),$$

where the first equality follows from Lemma E.4 (iii). Hence $B^j \subseteq B^i$.

Step 3. Fix $i, j \in I$ and assume $B_i^j \neq \emptyset$. Notice, since each message is sent with positive probability, for each $h_i \in B_i^j \subseteq M_i$ there is some $(\theta, y, h_{-i}) \in \Theta \times \overline{Y} \times M_{-i}$ such that $\phi(\theta, y, h_i, h_{-i}) > 0$. Notice, by Lemma E.4 (iii), $(h_i, h_{-i}) \in B_i^j \times B_{-i}^j$. Moreover, by Step 2, $(h_i, h_{-i}) \in B_i^k \times B_{-i}^k$ for each $k \in I$. Thus, by Step 1, $y \in W_i$ for each $i \in I$. Since at most one agent wins the object, we obtain a contradiction. We conclude that $\mathcal{M}$ conceals each E_i.

Now we show (i). Notice Lemma E.3 implies that $\mathcal{M}$ is BIC and IR. So, $\mathcal{M} \in \text{IMP}$. Lemma E.5 implies that $\text{VW}(\mathcal{M}) = \sup_{\mathcal{M}' \in \text{IMP}} \text{VW}(\mathcal{M}')$. Since $\mathcal{M}$ conceals each E_i, Lemma E.10 implies that $\text{PW}(\mathcal{M}) = \sup_{\mathcal{M}' \in \text{IMP}} \text{PW}(\mathcal{M})$. Therefore, $\mathcal{M}$ is revenue maximizing. (See Corollary 6.1.)

Finally, Part (ii) follows from Lemma E.5 and Part (iii) follows from Lemma E.10. □

Proofs of Propositions 6.2 and 6.3. Lemma 1 in Bergemann and Pesendorfer (2007) characterizes revenue maximizing auctions for the case of standard preferences. The rest of the cases follow from Lemmata E.7, E.9, and E.11. □

Proof of Lemma 6.1. Notice, the honest strategy in the Mechanism $(*)$ described in Section 2.1 is individually rational. Moreover, such mechanism provides expected revenue 6 to the auctioneer. This shows Part (i).

Fix $\mathcal{M} \in \text{IMP}$. To show Part (ii) it suffices to show that, $\text{VW}(\mathcal{M}) \leq 3$ and $\text{PW}(\mathcal{M}) \leq 3$. To show $\text{VW}(\mathcal{M}) \leq 3$, fix $\mathcal{M} \in \text{IMP}$. Notice that $v_A(0) = -6$, $v_A(6) = 6$. Hence,

$$\text{VW}(\mathcal{M}) = \sum_{\theta_A \in \Theta_A} Q_A(\theta_A) \cdot v(\theta_A) \cdot \mu_A(\theta_A) \leq 6 \cdot \mu_A(6) = 3.$$

To show $\text{PW}(\mathcal{M}) \leq 3$, write $\overline{M}_A := \{h_A \in M_A : f_A(h_A) = 4\}$ and $\underline{M}_A := M_A \setminus \overline{M}_A$. So, $\text{PW}(\mathcal{M}) = 4 \cdot \text{marg}_{M_A} \phi(\overline{M}_A)$. Thus, it suffices to show that $\text{marg}_{M_A} \phi(\overline{M}_A) \leq \frac{3}{4}$. To show this, observe that $\mathbb{E}_A[\mathbb{E}_B[\theta_A] \mid h_A] \geq 0$ for each $h_A \in \underline{M}_A$ and $\mathbb{E}_A[\mathbb{E}_B[\theta_A] \mid h_A] \geq 4$ for each $h_A \in \overline{M}_A$. Hence,

$$\sum_{h_A \in \overline{M}_A} 4 \cdot \text{marg}_{M_A} \phi(h_A) \leq \sum_{h_A \in M_A} \mathbb{E}_A[\mathbb{E}_B[\theta_A] \mid h_A] \cdot \text{marg}_{M_A} \phi(h_A) = \sum_{\theta_A \in \Theta_A} \mu_A(h_A) = 3,$$

where the first equality follows from Lemma E.4 (v). Therefore, $\text{marg}_{M_A} \phi(\overline{M}_A) \leq \frac{3}{4}$.

Finally, Part (iii) follows from Lemmata B.1 and B.2. □

Appendix F. Reduced-form utility functions

Fix an environment (I, Θ, Y) and $\hat{h}_i = (\hat{h}_i^k)_{k \in \mathbb{N}}, \widetilde{h}_i = (\widetilde{h}_i^k)_{k \in \mathbb{N}} \in H_i$. Let $\lambda \in [0, 1]$ and write $\lambda \cdot \hat{h}_i + (1 - \lambda) \cdot \widetilde{h}_i$ for the sequence of convex combinations $(\overline{h}_i^k)_{k \in \mathbb{N}}$ such that $\overline{h}_i^k = \lambda \cdot \hat{h}_i^k + (1 - \lambda) \cdot \widetilde{h}_i^k$ for each $k \in \mathbb{N}$.

Lemma F.1. *If $\overline{h}_i = \lambda \cdot \hat{h}_i + (1 - \lambda) \cdot \widetilde{h}_i$, then $\overline{h}_i \in H_i$ and $\overline{h}_i^\infty = \lambda \cdot \hat{h}_i^\infty + (1 - \lambda) \cdot \widetilde{h}_i^\infty$.*

Proof. The proof follows from the additive properties of measures and Kolmogorov's extension theorem. □

The utility function $u_i : \Theta_i \times Y \times H_i \to \mathbb{R}$ is **convex in** H_i, if for each $(\theta_i, y) \in \Theta_i \times Y$, $\lambda \in [0, 1]$, and $\hat{h}_i, \widetilde{h}_i \in H_i$, $u_i(\theta_i, y, \lambda \hat{h}_i + (1 - \lambda)\widetilde{h}_i) \leq \lambda u_i(\theta_i, y, \hat{h}_i) + (1 - \lambda) u_i(\theta_i, y, \widetilde{h}_i)$.

Lemma F.2. *The utility function $u_i : \Theta \times Y \times H_i \to \mathbb{R}$ representing unsophisticated image concerns (Example 2.5) is not convex in H_i.*

Proof. Fix $b > 0$ and set $E_i = \{\theta_i \in \Theta_i : \theta_i < b\}$. Fix $\hat{h}_i \in \mathcal{B}_i(E_i)$, $\widetilde{h}_i \notin \mathcal{B}_i(E_i)$ and set $\overline{h}_i = \frac{1}{2}\hat{h}_i + \frac{1}{2}\widetilde{h}_i$. Since $\widetilde{h}_i \notin \mathcal{B}_i(E_i)$, $\operatorname{marg}_{H_{-i}} \widetilde{h}_i^{\infty}(\mathcal{B}_{-i}(E_i)) < 1$. So, $\operatorname{marg}_{H_{-i}} \overline{h}_i^{\infty}(\mathcal{B}_{-i}(E_i)) < 1$ which implies $\overline{h}_i \notin \mathcal{B}_i(E_i)$. Consequently, $f_i(\overline{h}_i) = f_i(\widetilde{h}_i) = a > 0$ and $f_i(\hat{h}_i) = 0$. Therefore, $u_i(\cdot,\cdot,\widetilde{h}_i) = u_i(\cdot,\cdot,\overline{h}_i) > u_i(\cdot,\cdot,\hat{h}_i)$. So u_i is not convex in H_i. □

Proof of Proposition 7.1. Suppose that u_i is a reduced-form utility function of a Bayesian equilibrium $(\alpha_i)_{i\in I}$ of a game $\mathcal{G}$. Fix $\theta_i \in \Theta_i$ and $y \in Y$ and $h_i \in H_i$. Notice that

$$\begin{aligned}
u_i(\theta_i, y, h_i) &= \int_{A_i} \hat{v}_i(\theta_i, y, a_i \mid h_i, \alpha_{-i})\, d\alpha_i(\theta_i, y, h_i) \\
&= \sup_{a_i \in A_i} \hat{v}_i(\theta_i, y, a_i \mid h_i, \alpha_{-i}) \\
&= \sup_{a_i \in A_i} \int_{\Theta_{-i}\times H_{-i}} \int_{A_{-i}} v_i(\theta_i, y, \theta_{-i}, a_i, a_{-i})\, d\alpha_{-i}(\theta_{-i}, y, h_{-i})\, dh_i^{\infty},
\end{aligned}$$

where the first equality follows from the fact that $(\alpha_i)_{i\in I}$ is a Bayesian Equilibrium. Fix $\hat{h}_i, \widetilde{h}_i \in H_i$, and $\lambda \in [0,1]$. Write $\overline{h}_i = \lambda\hat{h}_i + (1-\lambda)\widetilde{h}_i$. By Lemma F.1, $\overline{h}_i \in H_i$ and $\overline{h}_i^{\infty} = \lambda\hat{h}_i^{\infty} + (1-\lambda)\widetilde{h}_i^{\infty}$. So,

$$\begin{aligned}
u_i(\theta_i, y, \overline{h}_i) &= \sup_{a_i \in A_i} \int_{\Theta_{-i}\times H_{-i}} \int_{A_{-i}} v_i(\theta_i, \theta_{-i}, y, a_i, a_{-i})\, d\alpha_{-i}(\theta_{-i}, y, h_{-i})\, d\overline{h}_i^{\infty} \\
&= \sup_{a_i \in A_i} \left[\lambda \int_{\Theta_{-i}\times H_{-i}} \int_{A_{-i}} v_i(\theta_i, \theta_{-i}, y, a_i, a_{-i})\, d\alpha_{-i}(\theta_{-i}, y, h_{-i})\, d\hat{h}_i^{\infty} \right. \\
&\qquad \left. + (1-\lambda) \int_{\Theta_{-i}\times H_{-i}} \int_{A_{-i}} v_i(\theta_i, \theta_{-i}, y, a_i, a_{-i})\, d\alpha_{-i}(\theta_{-i}, y, h_{-i})\, d\widetilde{h}_i^{\infty} \right] \\
&\le \sup_{a_i \in A_i} \left[\lambda \int_{\Theta_{-i}\times H_{-i}} \int_{A_{-i}} v_i(\theta_i, y, \theta_{-i}, a_i, a_{-i})\, d\alpha_{-i}(\theta_{-i}, y, h_{-i})\, d\hat{h}_i^{\infty} \right] \\
&\qquad + \sup_{a_i \in A_i} \left[(1-\lambda) \int_{\Theta_{-i}\times H_{-i}} \int_{A_{-i}} v_i(\theta_i, \theta_{-i}, y, a_i, a_{-i})\, d\alpha_{-i}(\theta_{-i}, y, h_{-i})\, d\widetilde{h}_i^{\infty} \right] \\
&= \lambda \cdot u_i(\theta_i, y, \hat{h}_i) + (1-\lambda) \cdot u_i(\theta_i, y, \widetilde{h}_i).
\end{aligned}$$

Therefore, $u_i(\theta_i, y, h_i)$ is convex in H_i. □

References

Aliprantis, Charalambos D., Border, Kim C., 2006. Infinite Dimensional Analysis. Springer, Berlin, London.

Alpizar, Francisco, Carlsson, Fredrik, Johansson-Stenman, Olof, 2008. Anonymity, reciprocity, and conformity: evidence from voluntary contributions to a national park in Costa Rica. J. Public Econ. 92 (5–6), 1047–1060.

Battigalli, Pierpaolo, Dufwenberg, Martin, 2009. Dynamic psychological games. J. Econ. Theory 144 (1), 1–35.

Battigalli, Pierpaolo, Dufwenberg, Martin, 2022. Belief-dependent motivations and psychological game theory. J. Econ. Lit. 60 (3), 833–882.

Battigalli, Pierpaolo, Generoso, Nicoló, 2021. Information flows and memory in games. Available at SSRN 4435785.

Battigalli, Pierpaolo, Corrao, Roberto, Dufwenberg, Martin, 2019a. Incorporating belief-dependent motivation in games. J. Econ. Behav. Organ. 167, 185–218.

Battigalli, Pierpaolo, Dufwenberg, Martin, Smith, Alec, 2019b. Frustration, aggression, and anger in leader-follower games. Games Econ. Behav. 117, 15–39.

Bénabou, Roland, Tirole, Jean, 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (5), 1652–1678.

Bergemann, Dirk, Pesendorfer, Martin, 2007. Information structures in optimal auctions. J. Econ. Theory 137 (1), 580–609.

Bos, Olivier, Pollrich, Martin, 2022. Auctions with signaling bidders: Optimal design and information disclosure. Available at SSRN 4252493. Working paper.

Bos, Olivier, Truyts, Tom, 2021. Auctions with signaling concerns. J. Econ. Manag. Strategy 30 (2), 420–448.

Brandenburger, Adam, Dekel, Eddie, 1993. Hierarchies of beliefs and common knowledge. J. Econ. Theory 59 (1), 189–198.

Calzolari, Giacomo, Pavan, Alessandro, 2006a. Monopoly with resale. Rand J. Econ. 37 (2), 362–375.

Calzolari, Giacomo, Pavan, Alessandro, 2006b. On the optimality of privacy in sequential contracting. J. Econ. Theory 130 (1), 168–204.

Calzolari, Giacomo, Pavan, Alessandro, 2009. Sequential contracting with multiple principals. J. Econ. Theory 144 (2), 503–531.

Carlsson, Hans, Van Damme, Eric, 1993. Global games and equilibrium selection. Econometrica, 989–1018.

Crawford, Vincent P., 2021. Efficient mechanisms for level-k bilateral trading. Games Econ. Behav. 127, 80–101.

Dasgupta, Partha, Hammond, Peter, Maskin, Eric, 1979. The implementation of social choice rules: some general results on incentive compatibility. Rev. Econ. Stud. 46 (2), 185–216.

Dillenberger, David, Sadowski, Philipp, 2012. Ashamed to be selfish. Theor. Econ. 7 (1), 99–124.

Doval, Laura, Skreta, Vasiliki, 2022. Mechanism design with limited commitment. Econometrica 90 (4), 1463–1500.
Doval, Laura, Skreta, Vasiliki, 2023. Optimal mechanism for the sale of a durable good. Theor. Econ., forthcoming.
Dworczak, Piotr, 2020. Mechanism design with aftermarkets: cutoff mechanisms. Econometrica 88 (6), 2629–2661.
Dziuda, Wioletta, Gradwohl, Ronen, 2015. Achieving cooperation under privacy concerns. Am. Econ. J. Microecon. 7 (3), 142–173.
Gan, Tan, 2022. Gacha game: when prospect theory meets optimal pricing. Working paper. arXiv preprint. arXiv:2208.03602.
Geanakoplos, John, Pearce, David, Stacchetti, Ennio, 1989. Psychological games and sequential rationality. Games Econ. Behav. 1 (1), 60–79.
Gershkov, Alex, Moldovanu, Benny, Strack, Philipp, Zhang, Mengxi, 2023. Optimal insurance: Dual utility, random losses and adverse selection. Am. Econ. Rev. 113 (10), 2581–2614.
Gibbard, Allan, 1973. Manipulation of voting schemes: a general result. Econometrica, 587–601.
Giovannoni, Francesco, Makris, Miltiadis, 2014. Reputational bidding. Int. Econ. Rev. 55 (3), 693–710.
Green, Jerry, 1987. "Making book against oneself," the independence axiom, and nonlinear utility theory. Q. J. Econ. 102 (4), 785–796.
Gul, Faruk, Pesendorfer, Wolfgang, 2001. Temptation and self-control. Econometrica 69 (6), 1403–1435.
Köszegi, Botond, 2006. Ego utility, overconfidence, and task choice. J. Eur. Econ. Assoc. 4 (4), 673–707.
Köszegi, Botond, Rabin, Matthew, 2007. Reference-dependent risk attitudes. Am. Econ. Rev. 97 (4), 1047–1073.
Köszegi, Botond, Rabin, Matthew, 2009. Reference-dependent consumption plans. Am. Econ. Rev. 99 (3), 909–936.
Lipnowski, Elliot, Mathevet, Laurent, 2018. Disclosure to a psychological audience. Am. Econ. J. Microecon. 10 (4), 67–93.
Mathevet, Laurent, Perego, Jacopo, Taneva, Ina, 2020. On information design in games. J. Polit. Econ. 128 (4), 000.
Myerson, Roger B., 1979. Incentive compatibility and the bargaining problem. Econometrica, 61–73.
Myerson, Roger B., 1981. Optimal auction design. Math. Oper. Res. 6 (1), 58–73.
Myerson, Roger B., 1982. Optimal coordination mechanisms in generalized principal–agent problems. J. Math. Econ. 10 (1), 67–81.
Pai, Mallesh M., Roth, Aaron, 2013. Privacy and mechanism design. ACM SIGecom Exch. 12 (1), 8–29.
Rayo, Luis, 2013. Monopolistic signal provision. B. E. J. Theor. Econ. 13 (1), 27–58.
Rivera Mora, Ernesto, 2023. Neutral mechanisms: on the feasibility of information sharing. Working paper.
Saran, Rene, 2011. Menu-dependent preferences and revelation principle. J. Econ. Theory 146 (4), 1712–1720.
Sugaya, Takuo, Wolitzky, Alexander, 2021. The revelation principle in multistage games. Rev. Econ. Stud. 88 (3), 1503–1540.
Warner, Stanley L., 1965. Randomized response: a survey technique for eliminating evasive answer bias. J. Am. Stat. Assoc. 60 (309), 63–69.