Neutral Mechanisms: On the Feasibility of Information Sharing*

Ernesto Rivera Mora†

Job Market Paper‡

December 12, 2022

Abstract

The paper analyzes information sharing in neutral mechanisms when an informed party will face future interactions with an uninformed party. Neutral mechanisms are mechanisms that do not rely on (1) the provision of evidence, (2) conducting experiments, (3) verifying the state, or (4) changing the after-game (i.e., the available choices and payoffs of future interactions). They include cheap talk, long cheap talk, noisy communication, mediation, money burning, and transfer schemes, among other mechanisms. To address this question, the paper develops a reduced-form approach that characterizes the agents' payoffs in terms of belief-based utilities. This effectively induces a psychological game, where the psychological preferences summarize information-sharing incentives. The first main result states that if an expert's reduced form (i.e., belief-based utility) satisfies a weak supermodularity condition between the state and hierarchies of beliefs, then there is a neutral mechanism that induces complete revelation of the state. Moreover, it identifies a mechanism that is easy to implement. The second main result states that if the expert's reduced-form representation (i.e., set of belief-based utilities) satisfies a strict submodularity condition between the state and the hierarchies of beliefs, neutral mechanisms are futile for any (relevant) information sharing. This implies a limit in the ability to use neutral mechanisms for information sharing. The paper goes on to show how the approach is useful in applications related to political economy and industrial organization.

*I thank Amanda Friedenberg, Andreas Blume, Inga Deimen, Stanley Reynolds, Derek Lemoine, Martin Dufwenberg, Ina Taneva, and Tiemen Woutersen for their helpful comments.

†Department of Economics, University of Arizona, eriveram [at] arizona.edu.

‡The current version of the paper can be found at www.ernestoriveramora.com/s/JMP.pdf.

1 Introduction

Throughout history, adverse outcomes have arisen from the concealment of information. An important example is firms that conceal information related to the health risks of their products. For instance, the tobacco industry took steps to conceal the health consequences of cigarettes from the public [Glantz, Slade, Bero, Hanauer, and Barnes, 1998]. Decades later, Purdue Pharma concealed the addictive properties of opioid-based painkillers, planting the seeds for the so-called "opioid epidemic" [Meier, 2018]. More recently, social networks have been accused of burying internal research related to the negative mental health effects they generate in young users [Wells, Horwitz, and Seetharaman, 2021]. These recurring negative outcomes raise important questions about the introduction of new products. Can policymakers create mechanisms that induce firms to reveal the harmful properties of their products? What kind of mechanisms can induce information sharing? What kind of mechanisms are doomed to fail? This paper studies the extent to which a broad class of mechanisms can or cannot induce an informed party to share its private information.

Understanding the possibility for or impossibility of information sharing is important beyond the extent to which the public or regulators can learn the health risks of products. For instance, the question also arises in antitrust regulation. There is reason to think that—even in the absence of collusion—information sharing can be harmful [Vives, 1984]. This might suggest that regulators should look for the presence of information sharing, whether or not collusion is present. To the extent that there are markets that preclude firms or trade associations from sharing relevant information, regulators need not worry about information sharing. (The results in Section 7.2 will point to such markets.) Likewise, the question arises in the analysis of auctions. For instance, in spectrum auctions, bidders typically know each other and may well have an incentive to share their private information, even if they do not engage in stronger forms of collusion. Yet, the textbook analysis of these auctions implicitly assumes that there is no information sharing (absent full collusion). This raises the question of whether the private information assumption is reasonable or whether bidders would share information.

To address these questions, the paper focuses on a model with two agents: an expert and a layman. The expert knows a payoff-relevant state of the world and the layman does not. Absent the means to share information, the agents play a game where payoffs depend on actions and the state of the world. For instance, this game may involve a regulator deciding whether to ban or approve a potentially addictive product of a firm. The firm knows if the product is safe or addictive, but the regulator does not. A designer seeks to construct a mechanism where both agents interact and share information prior to playing the game. For instance, a legislator may seek to design legal institutions that incentivize the firm to share its information with the regulator. This paper explores whether it is feasible or infeasible for the designer to construct a mechanism where the expert reveals the state to the layman. In particular, it asks whether a broad class of *neutral mechanisms* allows for information sharing.

Neutral mechanisms are extensive-forms where agents may exchange messages, privately receive signals about the behavior of other agents, and exchange monetary transfers. Importantly, neutral mechanisms require four independence conditions. First, they must satisfy *structural independence*. That is, action sets, information sets and action correspondences are independent of the state. This implies that the expert has no actions that directly reveal her private information. It rules out hard evidence and disclosure [Milgrom, 1981, Grossman, 1981]. Second, they must satisfy *statistical independence*. That is, the likelihoods of chance moves do not directly depend on the state. The only information that the layman gets about the state comes through the expert's actions in the mechanism. This rules out Blackwell [1953] experiments that are used in the information design literature [Kamenica and Gentzkow, 2011, Rayo and Segal, 2010, Taneva, 2019, Bergemann and Morris, 2019]. Third, they must satisfy *outcome independence*. That is, the outcome mapping (the mapping from terminal nodes to outcomes) does not depend on the state. This rules out state-contingent transfers, where the set of transfers available depends on the state. As a consequence, Spence-style signaling [Spence, 1978] is ruled out. Fourth, they must satisfy *game independence*. That is, the mechanism does not change the *after-game*, i.e., the game that agents play after the mechanism concludes. The mechanism can only affect the behavior in the after-game insofar as information transmission affects the agents' posterior beliefs. So, while the agents can commit to follow the rules of the mechanism, they cannot commit to change the rules of the after-game. This rules out delegation [Dessein, 2002] and arbitration [Goltsman, Hörner, Pavlov, and Squintani, 2009]. Moreover, it rules out game-contingent transfers [Krishna and Morgan, 2008], where the transfers depend on the actions chosen in the after-game.

Neutral mechanisms are "neutral" in that they do not depend on the state or the after-game. There are many examples of neutral mechanisms. They include cheap talk [Crawford and Sobel, 1982], long cheap talk, [Aumann and Hart, 2003], noisy communication channels [Blume, Board, and Kawamura, 2007], mediation [Goltsman, Hörner, Pavlov, and Squintani, 2009], money burning [Ben-Porath and Dekel, 1992, Austen-Smith and Banks, 2000], and mechanism-contingent transfers (transfers that depend on the outcome of the mechanism) [Myerson, 1982, Krishna and Morgan, 2008], among others. Their popularity arises from the fact that they do not require the strong assumptions that other mechanisms do: In many environments, (1) the expert does not have verifiable evidence; (2) the designer does not have access to Blackwell experiments; (3) the designer cannot construct state-contingent transfers, because they cannot verify the state; (4a) the designer cannot force the agents to change the after-game; and (4b) writing a contract on behavior in the after-game may not be feasible.[1] Given the interest in neutral mechanisms and the difficulty of implementing non-neutral mechanisms, it is important to understand their limitations. When do such mechanisms allow for information sharing? When are stronger tools—non-neutral mechanisms—needed?

Characterizing whether information sharing is possible in a neutral mechanism requires analyzing a class of dynamic games of asymmetric information. Each dynamic game is composed of a neutral mechanism—where agents potentially share information—followed by the after-game. A standard technique for solving dynamic games of finite length is backward induction: The analyst

[1] Game-contingent transfers require payments after actions are chosen and, thus, credibility of such transfers requires a contract. However, behavior in the after-game may not be verifiable.

first characterizes behavior at the last move and proceeds backwards in the tree. However, in this setting, the dynamic game need not have observable actions. In particular, the information conveyed to the layman by the mechanism may not be commonly known, when the mechanism concludes. So, each mechanism may lead to a non-trivial Bayesian game that requires its own analysis. Hence, employing backward induction across this class of mechanisms is not trivial.

The paper takes an alternative approach. Understanding the value of a given mechanism requires understanding the information it will convey and, so, the beliefs the players will have in the after-game. The key insight is that, for the purpose of evaluating a given mechanism, it is not important to understand the behavior in the after-game, but instead to understand the value of the information it conveys. With this in mind, the paper takes a *reduced-form approach*: It uses belief-based utilities as an instrument to summarize how information impacts the agents' payoffs in the after-game. Then, it appends these belief-based utilities to the mechanism, formally defining a dynamic psychological game [Geanakoplos, Pearce, and Stacchetti, 1989, Battigalli and Dufwenberg, 2009]. In the psychological game, the agents interact in a neutral mechanism and obtain belief-based payoffs, but, they do not engage in the after-game. The paper characterizes the possibility or impossibility of information sharing, by linking equilibria of these psychological games to equilibria of the original mechanism design problem.

There are subtleties in implementing the reduced-form approach. First, the belief-based utilities will depend on the agents' hierarchies of beliefs about the state. Note, they cannot only depend on the agents' first-order beliefs since, after the mechanism concludes, the information conveyed to the layman may not be commonly known. Moreover, they cannot only rely on high but finite-order beliefs, since behavior in games of incomplete information is sensitive to higher-order beliefs [Rubinstein, 1989, Carlsson and Van Damme, 1993, Morris and Shin, 2001]. This is illustrated in the application of Section 7.2, where the reduced form depends on all hierarchies of beliefs. Second, if there are multiple equilibria, a single profile of belief-based utilities may not be sufficient to characterize equilibrium payoffs in the after-game. For this reason, we will look at a set of appropriate belief-based utility profiles, called a *reduced-form representation*.

The paper uses the reduced-form approach to provide two sharp results regarding information sharing. It does so by defining conditions on reduced forms, which capture supermodularity and submodularity properties between the state and the expert's hierarchy of beliefs. The supermodularity condition captures the idea that an expert that observes a high state has weakly higher incentives to "be perceived" as having observed a high state. By contrast, the submodularity condition captures the idea that an expert that observes a low state has strictly higher incentives to "be perceived" as having observed high state. (To the best of my knowledge, this is the first paper to introduce these conditions.) The first main theorem is a positive result. Loosely speaking, it states that complete information sharing is possible as long as the expert's reduced form satisfies the supermodularity condition. Intuitively, if an expert with a high state has a higher willingness to pay to be perceived high, then there is a transfer scheme that incentivizes the expert to truthfully report the state to the layman. Moreover, the result identifies a simple neutral mechanism that achieves this. The second main theorem is a negative result. Loosely speaking, it states that any relevant information sharing is infeasible if the expert's reduced-form representation satisfies the submodularity condition. Intuitively, if an expert with a low state has a strictly higher willingness to pay to be perceived high, then no relevant information (not even partial) can be transmitted. As a consequence, the agents behave as if they didn't interact in the mechanism at all.

In applications, it is often simple to construct reduced forms and to verify whether they satisfy the supermodularity and submodularity conditions. To illustrate this, the paper provides a characterization of information sharing in two applications. Section 7.1 analyzes a parametrized class of games where only the layman takes an action and Section 7.2 analyzes a class of games where both agents take actions. Each application shows how the parameters of the original after-game translate into the submodularity/supermodularity properties of the reduced forms. As a consequence, the translation provides a complete taxonomy of the set of parameters that allow for or preclude information sharing. Moreover, it provides economic insights about the extent to which the parameters influence (or do not impact) the agents' ability to engage in information sharing.

The first application involves a policymaker and a bureaucrat. The bureaucrat knows the state of the world. The policy is chosen by an uninformed policymaker. Both agents have quadratic payoffs that depend on the selected policy and the realized state. The application seeks to characterize the types of disagreement that allow for or preclude information sharing. The key insight is that the absolute level of disagreement is irrelevant for information sharing. What matters is how changes in the state affects the "direction" of the agents' preferred policies. Applying the main theorems, full revelation of the state is possible if and only if the agents face "directional agreement" i.e., if both agents agree about the direction of how the policy should change in terms of the state. (This is independent of how distant the agents' preferred policies are.) If agents face "directional disagreement," i.e., if the agents' preferred actions move in different directions as the state changes, then the submodularity condition is satisfied and the negative result applies: The bureaucrat has strong preferences to deceive the policymaker, and, as a consequence, no relevant information can be transmitted.

The second application analyzes the interaction between two firms competing in a duopoly market. This interaction is modeled as a symmetric quadratic game. For instance, the game can represent a model of price competition (leading to strategic complements) or a model of quantity competition (leading to strategic substitutes). One of the firms observes industry demand. (This could be because the firm purchased a forecast from a third party [Rivera Mora, 2021a], or for other reasons.) Should an antitrust agency worry that the oligopolies share information?[2] Or does the strategic interaction between firms prevent any information sharing from happening? Applying the main theorems yields two results: First, full information sharing is possible if the agents' actions are strategic complements, i.e., if firms compete in prices. Second, information sharing is infeasible

[2]Think of the firms as committing to a mechanism by way of a trade association. That is, the trade association can be seen as a designer.

if the agents' actions are strategic substitutes, i.e., if firms compete in quantities.[3] When firms compete in prices, a good-news firm (i.e., a firm with knowledge of a high-demand shock) has higher incentives to induce higher joint prices than the bad-news firm. Hence, the good-news firm has higher incentives to induce "optimistic beliefs" and the supermodularity condition is satisfied. However, when firms compete in quantities, the effects are reversed: The good-news firm has more incentives to corner the market by inducing the other firm to decrease its quantity produced. That is, the good-news firm has more incentives to induce "pessimistic beliefs" and hence, the submodularity condition is satisfied.

This paper follows a long tradition of studying information sharing. Typically, this is asked in the context of specific mechanisms. See Crawford and Sobel [1982], Aumann and Hart [2003], Blume, Board, and Kawamura [2007], Goltsman, Hörner, Pavlov, and Squintani [2009], Myerson [1982], Krishna and Morgan [2008], Austen-Smith and Banks [2000], Ziv [1993] for examples of neutral mechanisms and Milgrom [1981], Grossman [1981], Spence [1978], Dessein [2002], Kamenica and Gentzkow [2011], Rayo and Segal [2010], Taneva [2019], Bergemann and Morris [2019], for examples of non-neutral mechanisms.

Much of the literature focuses on a particular neutral mechanism and a particular after-game and shows that *only* partial information sharing is feasible. (That is, full information sharing is infeasible.) Notable exceptions are Ziv [1993], Ottaviani [2000], and Krishna and Morgan [2008] —which use mechanism-contingent transfer schemes—and Austen-Smith and Banks [2000] and Kartik [2007]—which use money burning—to induce full information sharing. Ottaviani [2000], Austen-Smith and Banks [2000], Kartik [2007], and Krishna and Morgan [2008] focus on environments where only the layman can choose an action and the agents' payoffs are supermodular in the action and state. (Their result is related to Application 7.1.) In Ziv [1993], the agents' payoffs are not supermodular in the action and state. (See the discussion in Footnote 3.) Notably, these papers do not have a result on the impossibility of information sharing. The impossibility result is novel and its proof is more subtle. In particular, it requires showing that, for *each* neutral mechanism and *each* equilibrium, no relevant information can be shared. This result is proved by using a revelation principle for psychological games [Rivera Mora, 2021b].

This paper fits into a small literature that uses belief-based utilities as an instrument to summarize payoffs of future interactions. Dworczak [2020] uses belief-based utilities to summarize the bidders' payoffs in auctions with aftermarkets. There, the belief-based utilities are used as a primitive that represents the aftermarket's payoffs in terms of the beliefs of the aftermarket's participants. Since the outcome of the mechanism is assumed to be public, first-order beliefs uniquely determine higher-order beliefs. Hence, in his setting, it suffices to assume that the belief-based utilities only

[3] While the two conclusions are similar to results in Vives [1984], Gal-Or [1985], and Raith [1996], the mechanisms studied in those papers are not neutral. In particular, they allow for verifiable signals (evidence). Moreover, they rule out transfers. By contrast, Ziv [1993] uses neutral mechanisms to share information about production costs (as opposed to demand). He shows that, when firms engage in quantity competition, firms can use mechanism-contingent transfers to reveal information about production costs. This interesting result does not contradict the impossibility of information sharing for quantity competition shown here, since the two papers focus on different sources of uncertainty.

depend on first-order beliefs. Morris [2001] and Ottaviani and Sørensen [2006] both use belief-based utilities to summarize an advisor's reputational payoff in future interactions. In Ottaviani and Sørensen [2006], the belief-based utility is exogenous while, in Morris [2001], it is endogenous (as here). Both papers study a class of (cheap talk) neutral mechanisms, in which the agents' first-order beliefs uniquely determine higher-order beliefs. So, in their setting, it suffices to assume that the belief-based utilities only depend on first-order beliefs. Unlike these previous papers, this paper studies information sharing in a wide variety of neutral mechanisms, including mechanisms where the outcome is not publicly observable (e.g. mediation). In these mechanisms, first-order beliefs do not determine higher-order beliefs. Because higher-order beliefs may be important in the after-game, the belief-based utilities depend on the full hierarchy of beliefs.

At the surface, the supermodularity condition might resemble the condition in Van Zandt and Vives [2007]. However, the conditions are quite different. Their supermodularity is defined between actions and a single parameter that captures both payoff and belief types. By contrast, here supermodularity and submodularity are defined between states (i.e. payoff types) and hierarchies. They provide an exogenous order on payoff-belief types, which in turn determines the order on hierarchies. While the order on payoff types (or states) is often given by the application, it is unclear how to interpret the order on belief types. By contrast, here, the order on hierarchies is inherited from the order on states.

2 Illustrative Example

A firm has developed a new painkiller. The painkiller can be safe (state $\overline{\theta}$) or addictive (state $\underline{\theta}$), where $\overline{\theta} > \underline{\theta}$. Ex ante, the likelihood of the painkiller being safe ($\overline{\theta}$) is $\mu(\overline{\theta}) < \frac{1}{2}$. A regulator is choosing whether to *approve* the painkiller for sale or *ban* the painkiller. Payoffs are common knowledge and given as follows:

	$\overline{\theta}$	$\underline{\theta}$
approve	$1, 1$	$c, 0$
ban	$0, 0$	$0, 1$

Figure 2.1. Payoffs of firm (first) and regulator (second)

So, the regulator wants to approve the painkiller if and only if it is safe, i.e., if the state is $\overline{\theta}$. The firm's profit is 0 if the regulator *bans*. Profit is normalized to 1, if the painkiller is *approved* and safe. If the painkiller is *approved* and addictive, profit is c. Note, c can be greater than 1, if the firm's profit increases when the painkiller is addictive. But, c can be less than 1 if the firm internalizes costs of providing an addictive substance to the population—e.g., the cost of future fines, reputation, and lawsuits.

The firm has private information about the safety of the painkiller. In particular, the firm knows whether the state is $\overline{\theta}$ or $\underline{\theta}$, but the regulator does not. Given the prior, the regulator would *ban* the painkiller. Presumably a policymaker (the designer) would want the regulator to make an informed decision. Can the policymaker construct a neutral mechanism that induces the firm to share her information with the regulator?

An implication of the main results will be the following characterization.

Possibility/Impossibility of Information Sharing:

(1) If $c \leq 1$, then complete information sharing is possible.

(2) If $c > 1$, then, independent of the neutral mechanism employed, the regulator chooses the uninformed decision.

When $c > 1$, information may be shared, but it will not affect the regulator's decision. In that case, the designer would have to look beyond neutral mechanisms to help the regulator.

In this simple example, the possibility/impossibility of information sharing can be shown directly. Instead of doing so, we will make use of the reduced-form approach described in the main text. (This will illustrate the main tools of the paper.) We will focus the discussion on a particular simple class of neutral mechanism: One in which the firm chooses a costly message that is publicly observed. Write m for such a message. The message comes at a cost $y(m) \in \mathbb{R}$ for the firm. (Note, the cost may in fact be a benefit.) The main text will consider general neutral mechanisms.

After the agents interact in the mechanism—that is, after the firm chooses its message and it is observed—the regulator updates his prior belief about the state. Write $p(m)$ for the (endogenous) probability that the regulator assigns to $\overline{\theta}$ (safe) after observing the message m. Because the message is publicly observed, the firm assigns probability one to the regulator assigning probability $p(m)$ to $\overline{\theta}$ (safe). So, the parameter $p(m)$ captures all the relevant information about hierarchies of beliefs.

The regulator's behavior is effectively determined by their first-order beliefs—i.e., beliefs about the safety of the painkiller after a message (m) has been observed ($p(m)$). The reduced-form approach directly references those beliefs. First, it characterizes the agents' equilibrium payoffs in terms of the beliefs the agents may have after they have updated. This reduces the original analysis to a *psychological game*, in which agents interact in the mechanism and then obtain their belief-based utilities.[4] The key idea is that, for any given mechanism, the associated psychological game can be used to characterize the distribution of beliefs that can arise in equilibrium. This will allow us to address whether information sharing is possible.

With this in mind, we begin by characterizing the agents' equilibrium payoffs in terms of beliefs. Note, *ban* is optimal if and only if $p(m) \in [0, \frac{1}{2}]$ and *approve* is optimal if and only if $p(m) \in [\frac{1}{2}, 1]$. This implies that the payoff the regulator can get as a function of p—i.e., the posterior that the state is $\overline{\theta}$—is

$$u_r(p) = \max\{1 - p, p\}.$$

[4] In the original analysis, the agents interact in the mechanism and then play a regulation after-game.

The firm's payoffs depends on the likelihood that the firm assigns to the the regulator choosing *approve*. Because this depends on the regulator's posterior, this likelihood can be summarized by a mapping $\text{ap} : [0,1] \to [0,1]$ with

$$\text{ap}(p) = \mathbb{1}[p > \tfrac{1}{2}] + x \cdot \mathbb{1}[p = \tfrac{1}{2}],$$

for some $x \in [0,1]$. (If the firm believes that the regulator assigns probability $p > \frac{1}{2}$ to $\overline{\theta}$, then the firm believes the regulator approves; if the firm believes that the regulator assigns probability $p < \frac{1}{2}$, then the firm believes the regulator *bans*.) So, the firm's belief-based utility function is

$$u_f(\theta, p) = \begin{cases} \text{ap}(p) & \text{if } \theta = \overline{\theta} \\ c \cdot \text{ap}(p) & \text{if } \theta = \underline{\theta}. \end{cases}$$

Any such pair of functions (u_f, u_r) defines a psychological game and is called a *reduced form* of the regulation after-game. Notice there are many reduced forms corresponding to different values of $x \in [0,1]$.

The main results speak to the possibility or impossibility of information sharing based on the reduced form of the informed party. Loosely:

Main Theorem:

(1) If u_f is weakly supermodular on $\{\underline{\theta}, \overline{\theta}\} \times \{0,1\}$, then complete information sharing is possible.

(2) If u_f satisfies a strict submodularity condition, then no (relevant) information sharing is feasible.

A key step is defining super- and submodularity in terms of the agents' hierarchies of beliefs. Doing so requires imposing an order on the hierarchies that is determined by the order on states. In this example, the order imposed by supermodularity (part (1)) corresponds to the standard notion, i.e.,

$$u_f(\overline{\theta}, 1) - u_f(\overline{\theta}, 0) \geq u_f(\underline{\theta}, 1) - u_f(\underline{\theta}, 0).$$

That is, the firm's benefit as being *perceived* as safe ($p(m) = 1$) over addictive ($p(m) = 0$) is higher when the product is safe ($\overline{\theta}$). Submodularity (part (2)) is more subtle and requires ordering all of the posterior beliefs, not just the posteriors where the firm is perceived as either safe ($p(m) = 1$) over addictive ($p(m) = 0$). The idea will be introduced bellow.

To understand the result intuitively note that, under supermodularity, the firm with the safe painkiller ($\overline{\theta}$) has (weakly) higher incentives to induce approval of the painkiller. So, the safe firm has a higher willingness to pay to be perceived as safe rather than addictive. As a consequence, complete information sharing is possible. On the other hand, under strict submodularity, the addictive firm has (strictly) higher incentives to be perceived as safe. In a sense, the regulation after-game induces strong incentives for the addictive firm to deceive the regulator. As a consequence, in each mechanism and each equilibrium, the regulator always takes the uninformed action.

The remainder of the argument illustrates how the main result relates to the key parameter of the model, c. Start with the positive result. Observe that u_f is weakly supermodular on $\{\underline{\theta}, \overline{\theta}\} \times \{0,1\}$

if and only if $c \leq 1$. As a consequence, there is a mechanism that induces complete information sharing. One such mechanism has the firm directly choosing one of two messages: a high message $\overline{m}$ or a low message $\underline{m}$. The cost of the high message, $y(\overline{m})$, is in $[c, 1]$ and the cost of the low message, $y(\underline{m})$, is 0. (Note, $y(\overline{m}) \in [c, 1]$ is only feasible when $c \leq 1$.) Because $y(\overline{m}) \in [c, 1]$, there is an equilibrium of the psychological game where the safe firm ($\overline{\theta}$) chooses $\overline{m}$ and the addictive firm ($\underline{\theta}$) chooses $\underline{m}$. (The fact that the incentive constraints can be satisfied follows from u_f being supermodular.) This induces posterior beliefs $p(\overline{m}) = 1$ and $p(\underline{m}) = 0$. That is, there is complete information sharing.

The negative result is more subtle. To show it, fix a mechanism with a set messages M and a cost function $y : M \to \mathbb{R}$. The negative result will follow from two equilibrium properties:

(i) The regulator is more weakly likely to *approve* after observing messages sent by the safe firm $\overline{\theta}$ vs. by the addictive firm $\underline{\theta}$.

(ii) If the state θ does not impact the likelihood of *approval* (via a message), then no information is shared and the regulator always *bans*.

So, it suffices to show that there is no equilibrium where the regulator is strictly more likely to *approve* after observing a message sent by the safe firm. That is, in any equilibrium where the safe firm ($\overline{\theta}$) selects a message $\overline{m}$ with positive probability and the addictive firm ($\underline{\theta}$) selects a message $\underline{m}$ with positive probability, it follows that $\text{ap}(p(\underline{m})) \geq \text{ap}(p(\overline{m}))$. If this is satisfied, the message cannot impact the likelihood of *approval* (by (i)), and hence there is no relevant information sharing (by (ii)).

To show this, assume, by contradiction, that $\text{ap}(p(\overline{m})) > \text{ap}(p(\underline{m}))$. Since $c > 1$,

$$\begin{aligned}
[u_f(\underline{\theta}, p(\overline{m})) - u_f(\underline{\theta}, p(\underline{m}))] - [y(\overline{m}) - y(\underline{m})] &= c \cdot [\text{ap}(p(\overline{m})) - \text{ap}(p(\underline{m}))] - [y(\overline{m}) - y(\underline{m})] \\
&> [\text{ap}(p(\overline{m})) - \text{ap}(p(\underline{m}))] - [y(\overline{m}) - y(\underline{m})] \\
&= [u_f(\overline{\theta}, p(\overline{m})) - u_f(\overline{\theta}, p(\underline{m}))] - [y(\overline{m}) - y(\underline{m})] \\
&\geq 0, \qquad (1)
\end{aligned}$$

where the last equality follows from the fact that the safe firm ($\overline{\theta}$) chooses $\overline{m}$ in equilibrium. The strict inequality contradicts that the addictive firm ($\underline{\theta}$) chooses $\underline{m}$ in equilibrium. Therefore, $\text{ap}(p(\underline{m})) \geq \text{ap}(p(\overline{m}))$, as desired. Note, the strict inequality is a form of strict submodularity of the reduced form $u_f(\cdot, \cdot)$ requiring that

$$u_f(\underline{\theta}, p(\overline{m})) - u_f(\underline{\theta}, p(\underline{m})) > u_f(\overline{\theta}, p(\overline{m})) - u_f(\overline{\theta}, p(\underline{m})),$$

whenever $\text{ap}(p(\overline{m})) > \text{ap}(p(\underline{m}))$. The addictive firm ($\underline{\theta}$) has a strictly higher marginal benefit of approval, and hence a strictly higher incentive to induce a high posterior $p(m)$.

The no information-sharing result does not rely on the specific class of neutral mechanisms the discussion focused on: mechanisms that have publicly observed costly messages. The result holds for all neutral mechanisms. In particular, it allows for mechanisms that garble the firm's behavior through noise and mechanisms that make use of mediation schemes. In those cases, the firm will not directly select a signal that is publicly observed (such as a message). Instead, the regulator will observe signals provided by the mechanism. The firm will influence the distribution of signals through its behavior. Notice, since $u_f(\theta, p)$ is linear in $\mathrm{ap}(p)$, the argument above applies. The addictive firm is the one who has a (strictly) higher willingness to pay for high-posterior signals. Hence, it selects a signal distribution with a weakly higher expected value of $\mathrm{ap}(p)$ than the distribution selected by the safe firm. However, under Bayesian updating, signals with a higher value of $\mathrm{ap}(p)$ are more likely to come from the safe firm (by (i)). As a consequence, the state does not impact likelihood of *approval* and no information can be shared (by (ii)).

Remarkably, the feasibility of information sharing is, in a sense, discontinuous in the parameter c. Small changes in c lead to big changes in what information can be transmitted. It suggests that, when c is close to 1, a designer may want to find ways to lower c; doing so would involve using a non-neutral mechanism. This sharp discontinuity on the set of parameters also appears in the applications of Section 7.

3 Model

There are two agents: an expert (e) and a layman (ℓ). Write $i \in \{e, \ell\}$ for an agent and $-i$ for the agent in $\{e, \ell\} \backslash \{i\}$. The agents' payoffs depend on the state of the world. Let $\Theta \subsetneq \mathbb{R}$ be a finite set of states. The state is drawn from a common prior $\mu \in \Delta(\Theta)$ with full support. The expert observes the realization of the state and the layman does not. The agents then play a simultaneous move game. In that game, the set of actions for agent i is a metric space A_i. Write $A = A_e \times A_\ell$. The payoff function for agent i is a continuous mapping $\pi_i : \Theta \times A \to \mathbb{R}$. Write $G = ((A_i, \pi_i) : i \in \{e, \ell\})$ for that game. The game G is fixed throughout the analysis.

3.1 Neutral Mechanisms

A (neutral) mechanism is an extensive form, which is played after the expert learns the state but before the agents play the game of interest G. These mechanisms allow agents to exchange information by interacting in the mechanism and exchanging transfers. Because the mechanisms are neutral, they do not depend on the realized state and cannot change the game G. Thus, these mechanisms can be defined independently of both the realized state and G. This is the approach taken below.

Formally, a mechanism is described as follows: There is a finite set of nodes V with a precedence relation $\succsim$, such that $(V, \succsim)$ forms a tree. Write $\emptyset \in V$ for the root of the tree and $Z \subsetneq V$ for the set of terminal nodes. The terminal nodes in Z will correspond to the start of the game G.

In the extensive form, moves can be made by the expert, the layman, and chance (c). Each $i \in \{e, \ell, \mathrm{c}\}$ has an information partition on V, given by $\mathcal{I}_i \subseteq 2^V$. Note that $\mathcal{I}_i$ is a partition of both non-terminal and terminal nodes in the extensive form. The information partition $\mathcal{I}_i$ satisfies two conditions. First, it satisfies no absentmindedness, i.e., if $\{v, v'\} \subseteq I$ and $v \neq v'$, then is not the case that $v \succsim v'$ or that $v' \succsim v$. Second, the mechanism has an observable end, i.e., if $I_i \cap Z \neq \emptyset$, then $I_i \subseteq Z$. Call an information set $I_i \subseteq Z$ a **terminal information set**; write T_i for an arbitrary terminal information set. The set of terminal information sets for i is $\mathcal{T}_i \subseteq \mathcal{I}_i$.

Write X for the set of actions. Each $i \in \{e, \ell, \text{c}\}$ has an action correspondence $\mathcal{A}_i : \mathcal{I}_i \backslash \mathcal{T}_i \rightrightarrows X$ that specifies the actions that are available at each non-terminal information set. The information partition and the action correspondence $\mathcal{A}_i$ are such that agents have perfect recall. Chance's behavior is described (exogenously) by a behavioral strategy $\sigma_\text{c} \in \prod_{I_\text{c} \in \mathcal{I}_\text{c} \backslash \mathcal{T}_\text{c}} \Delta(\mathcal{A}_\text{c}(I_\text{c}))$. So σ_c describes the distribution of chance's actions at each $I_\text{c} \in \mathcal{I}_\text{c} \backslash \mathcal{T}_\text{c}$. The set of transfers for $i \in \{e, \ell\}$ is a finite set $Y_i \subsetneq \mathbb{R}$. Write $Y = Y_e \times Y_\ell$. The transfer function $\gamma_i : \mathcal{T}_i \rightarrow Y_i$ associates each terminal information set of i with a transfer that i receives. Notice that this implicitly assumes that i observes her transfer y_i. (Observe, the transfer depends on the mechanism but not in the state.)

The profile $\mathcal{M} = ((V, \succsim), X, (\mathcal{I}_i, \mathcal{A}_i : i \in \{e, \ell, \text{c}\}), \sigma_\text{c}, (Y_i, \gamma_i : i \in \{e, \ell\}))$ describes a **mechanism**. Note three features of the definition. First, because the mechanism is independent of the state, the set of nodes V does not contain information about the realization of the state. Second, the definition allows for explicit simultaneous moves. (This is captured by the fact that each node is in some information set of each agent i.) Nevertheless, the definition does not require simultaneous moves since the set of actions at a given information set can be a singleton. Third, no absentmindedness implies that each $i \in \{e, \ell, \text{c}\}$ knows the start of the game, i.e., $\{\emptyset\} \in \mathcal{I}_i$.

The definition of a mechanism captures the properties of a neutral mechanism. First, $\mathcal{M}$ is structurally independent. The tree, information sets, action set, and action correspondences do not depend on the realization of Θ. Second, $\mathcal{M}$ is statistically independent. The strategy of chance σ_c does not depend on the realization of Θ. Third, $\mathcal{M}$ is outcome independent. The set of transfers Y and the transfer functions do not depend on the realization of Θ. Fourth, $\mathcal{M}$ is game independent. The mechanism does not make reference to G and so cannot change G itself.

A given mechanism induces a set of pure strategies for i, $R_i = \prod_{I_i \in \mathcal{I}_i \backslash \mathcal{T}_i} \mathcal{A}_i(I_i)$. We think about elements of R_i as "reports" that i makes throughout the mechanism. Write $R = R_e \times R_\ell \times R_\text{c}$. Write $\psi : R \rightrightarrows V$ for the **path correspondence**. So, $\psi(r)$ denotes the set of nodes of V that constitute the path of play under r. Write $\zeta : R \times V \rightarrow Z$ for the **end node mapping**, where $\zeta(r, v) \in Z$ is the end node that would be realized if the game started at v and actions are subsequently played according to the strategy profile r. Say that $r \in R$ **allows** $I \subseteq V$ if $\psi(r) \cap I \neq \emptyset$ and say $r_i \in R_i$ **allows** $I \subseteq V$ if there is some (r_{-i}, r_c) so that $(r_i, r_{-i}, r_\text{c})$ allows I.

3.2 The Supergame

The mechanism $\mathcal{M}$ and the game G together induce a supergame, denoted by $(\mathcal{M}, G)$. The timing of the supergame is given as follows: Nature chooses state θ. The expert observes θ. The agents play $\mathcal{M}$. Each $i \in \{e, \ell\}$ observes a terminal information set $T_i \in \mathcal{T}_i$. Finally, agents play G. The payoffs of each agent i are quasilinear in the outcome of G and the transfer y_i. So, the payoff for i from $(\theta, a, y_i) \in \Theta \times A \times Y_i$ is $\pi_i(\theta, a) + y_i$.

To define the strategies in the supergame, it is useful to extend the action correspondence from $\mathcal{A}_i : \mathcal{I}_i \backslash \mathcal{T}_i \rightrightarrows X$ to $\mathcal{A}_i : \mathcal{I}_i \rightrightarrows X \cup A_e \cup A_\ell$ so that (i) for each $T_i \in \mathcal{T}_i$, $\mathcal{A}_i(T_i) = A_i$, and (ii) for each

$I_i \in \mathcal{I}_i \backslash \mathcal{T}_i$, $\mathcal{A}_i(I_i)$ corresponds to what it was originally in $\mathcal{M}$. A **behavioral strategy for the expert** is a mapping $\sigma_e : \Theta \to \prod_{I_e \in \mathcal{I}_e} \Delta(\mathcal{A}_e(I_e))$. A **behavioral strategy for the layman** is a vector $\sigma_\ell \in \prod_{I_\ell \in \mathcal{I}_\ell} \Delta(\mathcal{A}_\ell(I_\ell))$.

Each pair (θ, σ_e) induces a probability distribution over reports R_e given by

$$\mathbb{P}(r_e \mid \theta, \sigma_e) = \prod_{I_e \in \mathcal{I}_e \backslash \mathcal{T}_e} \sigma_e(\theta)(\text{proj}_{I_e} r_e).$$

Similarly, each σ_ℓ induces a probability distribution over reports R_ℓ given by

$$\mathbb{P}(r_\ell \mid \sigma_\ell) = \prod_{I_\ell \in \mathcal{I}_\ell \backslash \mathcal{T}_\ell} \sigma_\ell(\text{proj}_{I_\ell} r_\ell).$$

Likewise, write $\mathbb{P}(r_c \mid \sigma_c)$ for the probability distribution over R_c that σ_c induces.

3.3 Interim Belief Mappings

The interim belief mappings specify the beliefs that agents hold while interacting in the supergame. At each node, the expert (resp. the layman) has beliefs about which node has been reached (resp. which state was realized and which node has been reached).[5] Ultimately, the belief mappings will be endogenous.

Fix a mechanism $\mathcal{M}$. An **interim belief mapping for the expert** is a function $\beta_e : \Theta \times \mathcal{I}_e \to \Delta(V)$ such that, for each $(\theta, I_e) \in \Theta \times \mathcal{I}_e$, $\beta_e(\theta, I_e)(I_e) = 1$. Likewise, an **interim belief mapping for the layman** is a function $\beta_\ell : \mathcal{I}_\ell \to \Delta(\Theta \times V)$ such that, for each $I_\ell \in \mathcal{I}_\ell$, $\beta_\ell(I_\ell)(\Theta \times I_\ell) = 1$. Notice that, for each $(\theta, T_e, T_\ell) \in \Theta \times \mathcal{T}_e \times \mathcal{T}_\ell$, $\beta_e(\theta, T_e)(Z) = 1$ and $\beta_\ell(T_\ell)(Z) = 1$. So, at each terminal information set, both agents know that the mechanism has ended.

We will want the interim beliefs to be consistent with the strategy profile played. Toward that end, fix a strategy profile (σ_e, σ_ℓ) of the supergame $(\mathcal{M}, G)$. This strategy profile and the prior $\mu \in \Delta(\Theta)$ induce a distribution on $\Theta \times V$. For a given pair $(\theta, v) \in \Theta \times V$, the ex-ante probability that θ occurs and the path goes through v, given that (r_e, σ_ℓ) (resp. (σ_e, r_ℓ)) is played, is

$$\mathbb{P}(\theta, v \mid r_e, \sigma_\ell) = \sum_{(r_\ell, r_c) \in R_\ell \times R_c} \mu(\theta) \cdot \mathbb{P}(r_\ell \mid \sigma_\ell) \cdot \mathbb{P}(r_c \mid \sigma_c) \cdot \mathbb{1}[v \in \psi(r_e, r_\ell, r_c)], \quad \text{and}$$

$$\mathbb{P}(\theta, v \mid \sigma_e, r_\ell) = \sum_{(r_e, r_c) \in R_e \times R_c} \mu(\theta) \cdot \mathbb{P}(r_e \mid \theta, \sigma_e) \cdot \mathbb{P}(r_c \mid \sigma_c) \cdot \mathbb{1}[v \in \psi(r_e, r_\ell, r_c)].$$

Definition 3.1. *The interim beliefs $\beta = (\beta_e, \beta_\ell)$ are **consistent** with $\sigma = (\sigma_e, \sigma_\ell)$ if the following hold:*

(i) For each $r_e \in R_e$, $(\theta, I_e) \in \Theta \times \mathcal{I}_e$, and $v \in I_e$,

$$\beta_e(\theta, I_e)(v) \cdot \mathbb{P}(\{\theta\} \times I_e | r_e, \sigma_\ell) = \mathbb{P}(\theta, v | r_e, \sigma_\ell).$$

(ii) For each $r_\ell \in R_\ell$, $I_\ell \in \mathcal{I}_\ell$, and $(\theta, v) \in \Theta \times I_\ell$,

$$\beta_\ell(I_\ell)(\theta, v) \cdot \mathbb{P}(\Theta \times I_\ell | \sigma_e, r_\ell) = \mathbb{P}(\theta, v | \sigma_e, r_\ell).$$

[5] Recall that, since the mechanism satisfies structural independence, the realization of Θ is not part of the description of nodes V.

Consistency requires that, if I_e (resp. I_ℓ) is reached with positive probability under (r_e, σ_ℓ) (resp. (σ_ℓ, r_ℓ)), then e's (resp. ℓ's) beliefs about the elements of I_e (resp. $\Theta \times I_\ell$) are derived by the rule of conditional probability. Consistency imposes the implicit requirement that interim beliefs satisfy *own-action independence:* The probability that i assigns to each v (resp. each (θ, v)) at information set I_i is independent of the reporting strategy r_i that is used (provided that i uses an r_i that allows for I_i).[6] So, if β is consistent with (σ_i, σ_{-i}) and i deviates from σ_i, i still believes that $-i$ is playing accordingly to σ_{-i}.

3.4 Equilibrium

Fix $(\theta, T_e, T_\ell) \in \Theta \times \mathcal{T}_e \times \mathcal{T}_\ell$ such that $T_e \cap T_\ell \neq \emptyset$. The expected payoff from G given the strategy profile $\sigma = (\sigma_e, \sigma_\ell)$ and profile (θ, T_e, T_ℓ) is

$$\Pi_i(\sigma \mid \theta, T_e, T_\ell) = \int_{A_\ell} \int_{A_e} \pi_i(\theta, a_e, a_\ell) \; d\sigma_e(\theta, T_e) \; d\sigma_\ell(T_\ell).$$

So, the expert's expected payoff of σ given (θ, T_e) and β_e is

$$\Pi_e(\sigma \mid \theta, T_e, \beta_e) = \sum_{T_\ell \in \mathcal{T}_\ell} \Pi_e(\sigma \mid \theta, T_e, T_\ell) \cdot \beta_e(\theta, T_e)(T_\ell),$$

and the layman's expected payoff of σ given T_ℓ and β_ℓ is

$$\Pi_\ell(\sigma \mid T_\ell, \beta_\ell) = \sum_{(\theta, T_e) \in \Theta \times \mathcal{T}_e} \Pi_\ell(\sigma \mid \theta, T_e, T_\ell) \cdot \beta_\ell(T_\ell)(\{\theta\} \times T_e).$$

To talk about sequential rationality we need to compute the agent's payoffs at each interim information set. For this, note that the probability that agent i reaches T_i, given that I_i is reached and σ is played, is given by

$$\mathbb{P}(T_e \mid \theta, I_e, \sigma, \beta_e) = \sum_{v \in I_e} \sum_{r \in R} \beta_e(\theta, I_e)(v) \cdot \mathbb{P}(r \mid \theta, \sigma) \cdot \mathbb{1}[\zeta(v, r) \in T_e], \text{ and}$$

$$\mathbb{P}(T_\ell \mid I_\ell, \sigma, \beta_\ell) = \sum_{(\theta, v) \in \Theta \times I_\ell} \sum_{r \in R} \beta_\ell(I_\ell)(\theta, v) \cdot \mathbb{P}(r \mid \theta, \sigma) \cdot \mathbb{1}[\zeta(v, r) \in T_\ell].$$

So, the agents' interim payoffs at information sets I_e and I_ℓ are

$$U_e(\sigma \mid \theta, I_e, \beta_e) = \sum_{T_e \in \mathcal{T}_e} [\Pi_e(\sigma \mid \theta, T_e, \beta_e) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid \sigma, \theta, I_e, \beta_e), \text{ and}$$

$$U_\ell(\sigma \mid I_\ell, \beta_\ell) = \sum_{T_\ell \in \mathcal{T}_\ell} [\Pi_\ell(\sigma \mid T_\ell, \beta_\ell) + \gamma_\ell(T_\ell)] \cdot \mathbb{P}(T_\ell \mid \sigma, I_\ell, \beta_\ell).$$

Note that the interim payoffs include the expected payoffs from G and the transfers from the mechanism $\mathcal{M}$.

Definition 3.2. *An assessment $(\sigma_e, \sigma_\ell, \beta_e, \beta_\ell)$ satisfies* ***sequential rationality*** *if the following hold:*

(i) For each $(\theta, I_e) \in \Theta \times \mathcal{I}_e$ and σ'_e, $U_e(\sigma_e, \sigma_\ell \mid \theta, I_e, \beta_e) \geq U_e(\sigma'_e, \sigma_\ell \mid \theta, I_e, \beta_e)$.

(ii) For each $I_\ell \in \mathcal{I}_\ell$ and σ'_ℓ, $U_\ell(\sigma_e, \sigma_\ell \mid I_\ell, \beta_\ell) \geq U_\ell(\sigma_e, \sigma'_\ell \mid I_\ell, \beta_\ell)$.

[6] This terminology follows [Battigalli, Catonini, and De Vito, 2022].

If (σ, β) satisfies sequential rationality, then each agent is optimizing at each information set. Notice that optimization is required both within the mechanism $\mathcal{M}$ (at each non-terminal information set in $\mathcal{M}$) and within G (at each terminal information set of $\mathcal{M}$).

Definition 3.3. *Call* (σ, β) *a* ***perfect Bayesian equilibrium (PBE)*** *if the assessment* (σ, β) *satisfies sequential rationality and the belief mappings* β *are consistent with* σ*.*

A PBE requires that agents are optimizing at each information set and that beliefs are computed by the law of conditional probability whenever possible. We will focus on PBE that satisfy individually rationality. Agents may refuse to play $(\mathcal{M}, G)$ and get an exogenous **outside option**. The outside option of the expert is given by a state-dependent mapping $\underline{\pi}_e : \Theta \to \mathbb{R} \cup \{-\infty\}$. The outside option of the layman is given by $\underline{\pi}_\ell \in \mathbb{R} \cup \{-\infty\}$.[7]

Definition 3.4. *A perfect Bayesian equilibrium* (σ, β) *is* ***individually rational*** *if (1) for each state* $\theta \in \Theta$*,* $U_e(\sigma|\theta, \{\emptyset\}, \beta_e) \geq \underline{\pi}_e(\theta)$*, and (2)* $U_\ell(\sigma|\{\emptyset\}, \beta_\ell) \geq \underline{\pi}_\ell$*.*

Individual rationality requires that, after the expert learns the sate but before the mechanism is played, each agent's expected payoffs from a Bayesian equilibrium (σ, β) is higher than their outside option. (Recall that $\{\emptyset\}$ is the initial information set in the mechanism.)

4 A Reduced-Form Approach

This section introduces a novel technique to analyze mechanisms for information sharing. First, it analyzes the equilibrium behavior in Bayesian games associated with G. The beliefs in these Bayesian games correspond to the information the agents have after the expert has learned the state and the mechanism has been played. Second, it studies how behavior across these Bayesian games effectively induces preferences for information. Such preferences are modeled by utility functions that depend on the agents' hierarchies of beliefs—hence, defining a psychological game. Third, it shows how the equilibria of supergames are "equivalent" to the equilibria of a class of psychological games. Finally, it uses a version of the revelation principle for psychological games [Rivera Mora, 2021b] to capture all equilibria that arise from the relevant class of psychological games.

4.1 The Induced Bayesian Game

The first step is to understand how the information agents have after a mechanism $\mathcal{M}$ is played affects the equilibrium payoffs in G. Formally, the terminal information sets associated with $\mathcal{M}$ and behavior in $\mathcal{M}$ induce an information structure for the game G. These terminal information sets and associated belief mappings will induce a Bayesian game.

To define the induced Bayesian game, it will be useful to have two pieces of notation. First, note, from the perspective of the induced Bayesian game, the only important aspect of the mechanism

[7] Note, in practice, the layman's outside option can also depend on the state; since the layman does not know the state, $\underline{\pi}_\ell$ can be taken as the expected value.

is the set of terminal information sets. (They will impact beliefs in the associated Bayesian game.) With this in mind, it will be convenient to write $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$ for a mechanism where the sets of terminal information sets are $\mathcal{T}_e$ and $\mathcal{T}_\ell$. Second, interim beliefs will only be important if they can arise from the agents' updating in the mechanism. With this in mind, it will be convenient to write cons($\mathcal{M}$) for the set of interim belief mappings in $\mathcal{M}$ that are consistent (Definition 3.1) with some strategy profile. Notice that $\beta \in \text{cons}(\mathcal{M})$ implies that β is derived from a common prior.

Fix a supergame $(\mathcal{M}, G)$ with $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$ and interim belief mappings $\beta \in \text{cons}(\mathcal{M})$. After the agents finish playing $\mathcal{M}$, each agent i learns information associated with a terminal information set. Notice that the realized terminal information sets T_e and T_ℓ may not be singletons. So, the expert knows (θ, T_e) but may not know T_ℓ and the layman knows T_ℓ but may not know (θ, T_e). Thus, we can think of these information sets as reflecting types of the agents. Formally, the expert's set of types is $\Theta \times \mathcal{T}_e$ and the layman's set of types is $\mathcal{T}_\ell$. Since $\mathcal{M}$ has an observable end, each profile $(\theta, T_e, T_\ell) \in \Theta \times \mathcal{T}_e \times \mathcal{T}_\ell$ satisfies $\beta_e(\theta, T_e)(Z) = \beta_\ell(T_\ell)(\Theta \times Z) = 1$. Therefore, the probability that an expert of type (θ, T_e) assigns to T_ℓ is $\beta_e(\theta, T_e)(T_\ell)$ and the probability that a layman of type T_ℓ assigns to (θ, T_e) is $\beta_\ell(T_\ell)(\{\theta\} \times T_e)$.[8] Write BG$(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ for this **induced Bayesian game**.

Within this Bayesian game, the **expert's strategy** is a mapping $\hat{\sigma}_e : \Theta \times \mathcal{T}_e \to \Delta(A_e)$, and the **layman's strategy** is a mapping $\hat{\sigma}_\ell : \mathcal{T}_\ell \to \Delta(A_\ell)$. The agents' expected payoffs of the strategy profile $\hat{\sigma}$ given (θ, T_e, β_e) and (T_ℓ, β_ℓ) respectively are $\Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e)$ and $\Pi_\ell(\hat{\sigma} \mid T_\ell, \beta_\ell)$.

Definition 4.1. *Call the profile $\hat{\sigma} = (\hat{\sigma}_e, \hat{\sigma}_\ell)$ a **Bayesian equilibrium** of* BG$(\mathcal{T}_e, \mathcal{T}_\ell, \beta_e, \beta_\ell)$ *if the following hold:*

(i) For each $(\theta, T_e) \in \Theta \times \mathcal{T}_e$ and $\hat{\sigma}'_e$, $\Pi_e(\hat{\sigma}_e, \hat{\sigma}_\ell \mid \theta, T_e, \beta_e) \geq \Pi_e(\hat{\sigma}'_e, \hat{\sigma}_\ell \mid \theta, T_e, \beta_e)$.

(ii) For each $T_\ell \in \mathcal{T}_\ell$ and $\hat{\sigma}'_\ell$, $\Pi_\ell(\hat{\sigma}_e, \hat{\sigma}_\ell \mid T_\ell, \beta_\ell) \geq \Pi_\ell(\hat{\sigma}_e, \hat{\sigma}'_\ell \mid T_\ell, \beta_\ell)$.

4.2 Psychological Games

The reduced-form approach uses utility functions based on hierarchies of beliefs to summarize the equilibrium payoffs obtained in any induced Bayesian game. So, instead of thinking about payoffs from an equilibrium of $(\mathcal{M}, G)$, we suppress reference to G and set payoffs directly as a utility function that depends on the agent's hierarchies of beliefs. This effectively transforms the supergame into a psychological game.

To describe the psychological game, we will need to introduce hierarchies of beliefs. Let $D^1_\ell = \Theta$ and $D^1_e = \{\diamond\}$, where $\diamond$ is a trivial element. These represent the first-order domain of uncertainty for the layman and the expert. (Notice that the expert has no uncertainty, and so, her first-order domain of uncertainty is trivial.) The set of first-order beliefs of agent i is $H^1_i = \Delta(D^1_i)$.

Inductively define the sets D^k_i and H^k_i as follows: Assume the sets D^k_i and H^k_i are defined for k. Then $D^{k+1}_i = D^k_i \times H^k_{-i}$ is the $(k+1)$-order domain of uncertainty of agent i and

$$H^{k+1}_i = \left\{(\mu^1_i, ..., \mu^{k+1}_i) \in H^k_i \times \Delta(D^{k+1}_i) : \text{marg}_{D^k_i} \mu^{k+1}_i = \mu^k_i\right\}$$

[8] Formally, in a Bayesian game the agents priors are mappings $\hat{\beta}_e : \Theta \times \mathcal{T}_e \to \Delta(\mathcal{T}_\ell)$ and $\hat{\beta}_\ell : \mathcal{T}_\ell \to \Delta(\Theta \times \mathcal{T}_e)$. For notational simplicity, we use the notation of (β_e, β_ℓ) instead of the correct notation $(\hat{\beta}_e, \hat{\beta}_\ell)$.

is the set of collectively coherent $(k+1)$-order beliefs of agent i. Note that, if $(\mu_i^1, ..., \mu_i^{k+1}) \in H_i^{k+1}$, then $(\mu_i^1, ..., \mu_i^n) \in H_i^n$ for all $n \leq k$; that is, each $(\mu_i^1, ..., \mu_i^{k+1}) \in H_i^{k+1}$ is coherent.

Write

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

for the set of i's **collectively coherent hierarchies of beliefs**. So, $h_i = (\mu_i^1, \mu_i^2, ...) \in H_i$ is a particular hierarchy of beliefs for agent i. Notice that there are hierarchies of beliefs $\tilde{h}_e = (\tilde{\mu}_e^1, \tilde{\mu}_e^2, ..) \in H_e$ and $\tilde{h}_\ell = (\tilde{\mu}_\ell^1, \tilde{\mu}_\ell^2, ..) \in H_\ell$ which are induced by the common prior $\mu \in \Delta(\Theta)$. The common prior hierarchies are such that each k-th order belief $\tilde{\mu}_i^k$ has finite support.

The **belief structure** is $H = H_e \times H_\ell$. Each hierarchy $h_i \in H_i$ can be mapped to a belief on $D_i^1 \times H_{-i}$, so that the marginals coincide with that specified by h_i. In particular, there is a **canonical homeomorphism** $\eta_i : H_i \to \Delta(D_i^1 \times H_{-i})$ so that $\eta_i(h_i)$ is the canonical extension of h_i. (See [Mertens and Zamir, 1985] and [Brandenburger and Dekel, 1993].)

A **belief-based utility** for the expert (resp. the layman) is a bounded measurable function $u_e : \Theta \times H_e \to \mathbb{R}$ (resp. $u_\ell : H_\ell \to \mathbb{R}$).[9] A mechanism $\mathcal{M}$ and belief-based utilities (u_e, u_ℓ) induce a **psychological game** $(\mathcal{M}, u_e, u_\ell)$. The timing of the psychological game is as follows: Nature chooses the state θ, the expert observes θ, and the agents play $\mathcal{M}$. Each agent i observes a terminal information set $T_i \in \mathcal{T}_i$.

As in the supergame, the agent's interim belief mappings are $\beta_e : \Theta \times \mathcal{I}_e \to \Delta(V)$ and $\beta_\ell : \mathcal{I}_\ell \to \Delta(\Theta \times V)$. (Again, $\beta = (\beta_e, \beta_\ell)$ will be endogenous.) For each $(\theta, T_e) \in \Theta \times \mathcal{T}_\ell$ (resp. $T_\ell \in \mathcal{T}_\ell$), β induces a hierarchy $h_e \in H_e$ (resp. $h_\ell \in H_\ell$). Write $\delta_e : \Theta \times \mathcal{T}_e \to H_e$ and $\delta_\ell : \mathcal{T}_\ell \to H_\ell$ for the induced **hierarchy mappings**. (See Appendix A.) Notice that δ_e and δ_ℓ depend on β. For notational convenience, the reference to β is suppressed.

The payoffs in the psychological game are quasilinear in the psychological payoff u_i and the transfer y_i. So, if the state is θ, the expert observes T_e, and has hierarchy mapping δ_e, then the expert's payoffs are $u_e(\theta, \delta_e(\theta, T_e)) + \gamma_e(T_e)$. Likewise, if the layman observes T_ℓ and has hierarchy mapping δ_ℓ, then the layman's payoffs are $u_\ell(\delta_\ell(T_\ell)) + \gamma_\ell(T_\ell)$.

Notice that a behavioral strategy in the psychological game $(\mathcal{M}, u_e, u_\ell)$ differs from a behavioral strategy in the supergame $(\mathcal{M}, G)$ insofar as it does not need to specify how G is played. Hence, a behavioral strategy for the expert is a mapping $\rho_e : \Theta \to \prod_{I_e \in \mathcal{I}_e \setminus \mathcal{T}_e} \Delta(\mathcal{A}_e(I_e))$ and a behavioral strategy for the layman is a vector $\rho_\ell \in \prod_{I_\ell \in \mathcal{I}_\ell \setminus \mathcal{T}_\ell} \Delta(\mathcal{A}_\ell(I_\ell))$. So, ρ_i effectively mixes between reporting strategies R_i in $\mathcal{M}$.

As in the supergame, a strategy profile $\rho = (\rho_e, \rho_\ell)$ and the prior μ induce a distribution on $\Theta \times V$. Write $\mathbb{P}(\theta, v | r_e, \rho_\ell)$ (resp. $\mathbb{P}(\theta, v | \rho_e, r_\ell)$) for the ex-ante probability that θ occurs and the path goes through v given that (r_e, ρ_ℓ) (resp. (ρ_e, r_ℓ)) is played. Likewise, write $\mathbb{P}(T_e \mid \theta, I_e, \rho, \beta_e)$ (resp. $\mathbb{P}(T_\ell \mid I_\ell, \rho, \beta_\ell)$) for the probability of reaching T_e (resp. T_ℓ) given $(\theta, I_e, \rho, \beta_e)$ (resp. $(I_\ell, \rho, \beta_\ell)$).

[9] Notice, since the layman does not observe the state, his belief-based utility does not directly depend on Θ; instead it depends on his first-order beliefs about Θ (associated with the hierarchy h_ℓ).

Given $\beta = (\beta_e, \beta_\ell)$, the agents' interim expected payoffs at (θ, I_e) (resp. I_ℓ) are

$$\mathcal{U}_e(\rho \mid \theta, I_e, \beta) = \sum_{T_e \in \mathcal{T}_e} [u_e(\theta, \delta_e(\theta, T_e)) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid \theta, I_e, \rho, \beta_e), \text{ and}$$

$$\mathcal{U}_\ell(\rho \mid I_\ell, \beta) = \sum_{T_\ell \in \mathcal{T}_\ell} [u_\ell(\delta_\ell(T_\ell)) + \gamma_\ell(T_\ell)] \cdot \mathbb{P}(T_\ell \mid I_\ell, \rho, \beta_\ell).$$

Recall that U_i represents i's interim payoffs of the supergame $(\mathcal{M}, G)$, while $\mathcal{U}_i$ represents i's interim payoffs of the psychological game $(\mathcal{M}, u_e, u_\ell)$. Notice that while U_i depends only on β_i, $\mathcal{U}_i$ depends on both β_i and β_{-i} because both are inputs into i's hierarchy of beliefs.

Call (ρ, β) a **perfect Bayesian equilibrium (PBE)** of the psychological game $(\mathcal{M}, u_e, u_\ell)$ if the assessment (ρ, β) satisfies sequential rationality and the belief mappings β are consistent with ρ. It requires:

Definition 4.2. *The assessment $((\rho_e, \rho_\ell), \beta)$ satisfies **sequential rationality** in $(\mathcal{M}, u_e, u_\ell)$ if the following hold:*

(i) For each $(\theta, I_e) \in \Theta \times (\mathcal{I}_e \backslash \mathcal{T}_e)$ and ρ'_e, $\mathcal{U}_e(\rho_e, \rho_\ell \mid \theta, I_e, \beta) \geq \mathcal{U}_e(\rho'_e, \rho_\ell \mid \theta, I_e, \beta)$.

(ii) For each $I_\ell \in (\mathcal{I}_\ell \backslash \mathcal{T}_\ell)$ and ρ'_ℓ, $\mathcal{U}_\ell(\rho_e, \rho_\ell \mid I_\ell, \beta) \geq \mathcal{U}_\ell(\rho_e, \rho'_\ell \mid I_\ell, \beta)$.

Definition 4.3. *The interim beliefs β are **consistent** with ρ if the following fold:*

(i) For each $r_e \in R_e$, $(\theta, I_e) \in \Theta \times \mathcal{I}_e$, and $v \in I_e$,

$$\beta_e(\theta, I_e)(v) \cdot \mathbb{P}(\{\theta\} \times I_e | r_e, \rho_\ell) = \mathbb{P}(\theta, v | r_e, \rho_\ell).$$

(ii) For each $r_\ell \in R_\ell$, $I_\ell \in \mathcal{I}_\ell$ and $v \in I_\ell$,

$$\beta_\ell(I_\ell)(\theta, v) \cdot \mathbb{P}(\Theta \times I_\ell | \rho_e, r_\ell) = \mathbb{P}(\theta, v | \rho_e, r_\ell).$$

While the psychological game does not explicitly model actions in terminal information sets, the belief-based payoffs still require updating at terminal information sets. So, in contrast to the supergame, sequential rationality here does not require optimization at terminal information sets. But, as in the supergame, consistency is required at all information sets, including terminal information sets.

4.3 Reduced Forms

Reduced forms formalize the idea of using belief-based utility functions to capture the equilibrium payoffs in the induced Bayesian games.

Definition 4.4. *Call (u_e, u_ℓ) a **reduced form** for G if, for each mechanism $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$ and interim belief mappings $\beta = (\beta_e, \beta_\ell) \in \text{cons}(\mathcal{M})$ (that induce (δ_e, δ_ℓ)), there is a Bayesian equilibrium $\hat{\sigma}$ of $\text{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta_e, \beta_\ell)$ such that the following hold:*

(i) For each $(\theta, T_e) \in \Theta \times \mathcal{T}_e$, $u_e(\theta, \delta_e(\theta, T_e)) = \Pi_e(\hat{\sigma} | \theta, T_e, \beta_e)$.

(ii) For each $T_\ell \in \mathcal{T}_\ell$, $u_\ell(\delta_\ell(T_\ell)) = \Pi_\ell(\hat{\sigma} | T_\ell, \beta_\ell)$.

A pair of belief-based utility functions (u_e, u_ℓ) is a reduced form for G if it captures equilibrium payoffs of each induced Bayesian game, by only making reference to the agents' hierarchies of beliefs. So, effectively, (u_e, u_ℓ) captures preferences for information induced by G.

The goal of the reduced-form approach is to use equilibria of the induced psychological games to characterize the equilibria of supergames. To do so, note that there are two essential components of a PBE of a supergame $(\mathcal{M}, G)$: transfers exchanged in the mechanism and payoffs from the induced Bayesian game. Definition 4.4 refers to the later but not the former. With this in mind, fix an assessment (σ, β) and note that the expected transfers from (σ, β) are

$$\mathbb{E}Y_e(\theta|\mathcal{M}, \sigma) = \sum_{T_e \in \mathcal{T}_e} \gamma_e(T_e) \cdot \mathbb{P}(T_e|\sigma, \theta)$$

and

$$\mathbb{E}Y_\ell(\mathcal{M}, \sigma) = \sum_{T_\ell \in \mathcal{T}_\ell} \gamma_\ell(T_\ell) \cdot \mathbb{P}(T_\ell|\sigma).$$

Similarly, write $\mathbb{E}Y_e(\theta|\mathcal{M}, \rho)$ and $\mathbb{E}Y_\ell(\mathcal{M}, \rho)$ for the agents expected transfers form the profile (ρ, β) in psychological game $(\mathcal{M}, u_e, u_\ell)$.

Definition 4.5. *A PBE (σ, β) of the supergame $(\mathcal{M}, G)$ and a PBE (ρ, β') of the psychological game $(\mathcal{M}', u_e, u_\ell)$ are **equivalent** if the following hold:*

(i) For each $\theta \in \Theta$, $U_e(\sigma|\theta, \{\emptyset\}, \beta_e) = \mathcal{U}_e(\sigma|\theta, \{\emptyset\}, \beta')$ and $\mathbb{E}Y_e(\theta|\mathcal{M}, \sigma) = \mathbb{E}Y_e(\theta|\mathcal{M}', \rho')$.

(ii) $U_\ell(\sigma|\{\emptyset\}, \beta_e) = \mathcal{U}_\ell(\sigma|\{\emptyset\}, \beta')$ and $\mathbb{E}Y_\ell(\mathcal{M}, \sigma) = \mathbb{E}Y_e(\mathcal{M}', \rho)$.

Equivalence captures the idea that transfers and total payoffs are the same in the supergame and psychological game. Importantly, the definition allows equivalence of equilibria across the supergame and the psychological game, even if the mechanisms $\mathcal{M}$ and $\mathcal{M}'$ are different. The following lemma states that the psychological games induced by reduced-forms capture equivalent equilibria of a supergame.

Lemma 4.1. *Fix a psychological game $(\mathcal{M}, u_e, u_\ell)$ where (u_e, u_ℓ) is a reduced form of G. For each PBE (ρ, β) of $(\mathcal{M}, u_e, u_\ell)$, there is a strategy profile σ such that (σ, β) is a PBE of $(\mathcal{M}, G)$ that is equivalent to (ρ, β).*

Lemma 4.1 states that each equilibrium of an associated psychological game $(\mathcal{M}, u_e, u_\ell)$ induces an equivalent equilibrium of the supergame associated with $\mathcal{M}$. The lemma implies that, for a given reduced form, the set of equilibria across all associated psychological games capture a subset of equilibria across all supergames. The lemma is silent about whether a reduced form captures all equilibria across all supergames. If each induced Bayesian game has a unique equilibrium, then a reduced form (u_e, u_ℓ) indeed captures all equilibria. (See Lemma B.13.) Applications 7.1 and 7.2 are examples of this case. If, however, some induced Bayesian games have multiple equilibria, then a reduced form (u_e, u_ℓ) need not capture all equilibria. The example in Section 2 illustrates this case. At the posterior $p = \frac{1}{2}$, the regulator is indifferent between the two actions. Since only the regulator is active in the after-game, at the posterior $p = \frac{1}{2}$, each probability of *approval* $(x \in [0, 1])$

is associated with an equilibrium distribution of some induced Bayesian game. So, to capture all the equilibrium payoffs, we must consider all reduced forms (u_f, u_r) associated with some such x.

Definition 4.6. *Fix a set* RF *of reduced forms of* G. *Say* RF *is a* ***reduced-form representation*** *of* G *if, for each mechanism* $\mathcal{M}$ *and each PBE* (σ, β) *of* $(\mathcal{M}, G)$*, there is a mechanism* $\mathcal{M}'$*, a reduced form* $(u_e, u_\ell) \in \mathrm{RF}$*, and a PBE of* $(\mathcal{M}', (u_e, u_\ell))$ *that is equivalent to* (σ, β)*.*

A reduced-form representation RF of G characterizes the equilibria of supergames as equilibria of psychological games. The PBE of the class of supergames $\{(\mathcal{M}, G) : \mathcal{M} \text{ is a mechanism}\}$ are equivalent to the PBE of the class of psychological games $\{(\mathcal{M}, u_e, u_\ell) : \mathcal{M} \text{ is a mechanism}, (u_e, u_\ell) \in \mathrm{RF}\}$. Intuitively, different reduced forms in RF may capture differences in behaviour for fixed hierarchies of beliefs. So, by considering all $(u_e, u_\ell) \in \mathrm{RF}$, the class of psychological games effectively captures all equilibria. Importantly, RF need not include all the reduced forms of G to be a reduced form representation. Section 8.1 discusses the existence of reduced-form representation and provides tools for finding them.

4.4 Extended Direct Mechanisms for Psychological Games

Following Rivera Mora [2021b], to analyze the PBE that emerge in psychological games, we need only analyze a class of extended direct mechanisms. Extended direct mechanisms differ from the textbook formulation of direct mechanisms, in that the designer selects both a transfer and a "hierarchy-message" to send to each agent. In particular, under an extended direct mechanism, the expert reports a state (only) to the designer. Given the report, the designer selects a transfer and private message for each agent. Each agent i (only) observes their transfer and their message. Importantly, the message that each agent receives is a suggestion of the hierarchies of beliefs that the agent should have.

Fix finite sets $M_e \subsetneq H_e$ and $M_\ell \subsetneq H_\ell$. Anticipating that these sets will serve as sets of messages the mechanism will send to the agents (about the hierarchies they should hold) we refer to the elements of M_e and M_ℓ as hierarchy-messages. Say that $M_e \times M_\ell$ is **belief closed** if, for each $h_i \in M_i$, $\eta_i(h_i)(D_i^1 \times M_{-i}) = 1$. (Recall that $\eta_i : H_i \to \Delta(D_{-i}^1 \times H_{-i})$ is i's canonical homeomorphism.) That is, $M_e \times M_\ell$ is belief closed if each hierarchy-message in M_i only assigns positive probability to hierarchy-messages in M_{-i}.

An **extended direct mechanism**, $\mathcal{M}^d = (\Theta, (Y_i, M_i : i \in \{e, \ell\}), m)$, is defined as follows: The set $Y_i \subsetneq \mathbb{R}$ is a finite set of transfers for agent i. The set $M_i \subsetneq H_i$ is a finite set of private hierarchy-messages for agent i, so that $M = M_e \times M_\ell$ is belief closed. The mapping $m : \Theta \to \Delta(Y \times M)$ is a **protocol** that describes the likelihood of chance selecting transfers and hierarchy-messages, given each report.

Notice $\mathcal{M}^d$ is a neutral mechanism in the sense of Section 3.1: The expert reports a state $\theta \in \Theta$ and chance selects $(y, h) \in Y \times M$ according to the distribution $m(\theta)$. Hence, the set of terminal nodes is $Z = \Theta \times Y \times M$, where the set Θ represents the reported state.[10] The expert's terminal

[10] Notice that the reported state may be different than the real state.

information sets are of the form $I_e = \{\theta\} \times \{y_e\} \times Y_\ell \times \{h_e\} \times M_\ell$ and the layman's information sets are of the form $I_\ell = \Theta \times Y_e \times \{y_\ell\} \times M_e \times \{h_\ell\}$.

The set of reporting strategies for the expert is $R_e = \Theta$, and the set of reporting strategies for the layman is $R_\ell = \{\diamond\}$, where $\diamond$ is a trivial action. So, in the associated psychological game, a strategy for the expert maps Θ to reporting strategies in $R_e = \Theta$. Write $\rho_e^* : \Theta \to \Delta(R_e)$ for the expert's **honest strategy**. i.e., for the strategy with $\rho_e^*(\theta)(\theta) = 1$ for each $\theta \in \Theta$. Under the **honest strategy profile** $\rho^* = (\rho_e^*, \rho_\ell^*)$, the expert truthfully reports the state. (Note that ρ_ℓ^* is trivial.) Write $\beta^* = (\beta_e^*, \beta_\ell^*)$ for **honest interim belief mappings**, i.e., belief mappings that are consistent with the honest strategy profile.

An extended direct mechanism $\mathcal{M}^d$ and the honest profile (ρ^*, β^*) induce an **ex-ante probability measure** $\phi \in \Delta(\Theta \times Y \times M)$, defined by $\phi(\theta, y, h) = \mu(\theta) \cdot m(\theta)(y, h)$.[11] This measure is the ex-ante distribution of terminal nodes (i.e., reports, transfers, and hierarchy-messages) that arise under truth-telling in $\mathcal{M}^d$.

Notice that different extended direct mechanisms may differ in what information is shared. In particular, the layman may learn the state in some mechanism but not in others. Moreover, in principle, the message that i receives may not coincide with the actual hierarchy that i has. However, for a particular class of mechanisms, namely "credible mechanisms," in equilibrium, the actual hierarchies coincide with the hierarchy-messages the agents receive.

Definition 4.7. *An extended direct mechanism $\mathcal{M}^d$ is **credible** if, for each $(\theta, y, h) \in \Theta \times Y \times M$, the following hold:*

(i) $\eta_\ell(h_\ell)(\theta, h_e) \cdot \text{marg}_{Y_\ell \times M_\ell} \phi(y_\ell, h_\ell) = \text{marg}_{\Theta \times Y_\ell \times M} \phi(\theta, y_\ell, h_e, h_\ell)$.

(ii) $\eta_e(h_e)(h_\ell) \cdot \text{marg}_{\Theta \times Y_e \times M_e} \phi(\theta, y_e, h_e) = \text{marg}_{\Theta \times Y_e \times M} \phi(\theta, y_e, h_e, h_\ell)$.

Recall that $\eta_i(h_i)$ is the canonical extension of h_i. Notice that the layman updates his beliefs about (θ, h_e) based on his hierarchy-message h_ℓ and the transfer y_ℓ he receives. Credibility requires that the layman's message h_ℓ is effectively derived by computing $\phi(\theta, h_e, y_\ell, h_\ell)$ conditional on (y_ℓ, h_ℓ). (Formally, the canonical extension of h_ℓ is derived in that way.) Notice, this requires that this conditional is constant on y_ℓ (whenever (y_ℓ, h_ℓ) has positive probability). A consequence of credibility is that the agents' posterior beliefs coincide with the hierarchy-messages they receive at all information sets that are reachable under the honest strategy profile. (See Lemma 4.1 in Rivera Mora [2021b].) So, for any pair of transfers and hierarchy-messages that the protocol can send, the updated belief only depends on the hierarchy-message, i.e., the transfer provides no further information.

Proposition 4.1. *Belief-based utilities (u_e, u_ℓ) are a reduced-form of G if and only if, for each credible direct mechanism $\mathcal{M}^d = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$ and honest beliefs $\beta^* = (\beta_e^*, \beta_\ell^*)$ thereof, there is a strategy profile $\hat{\sigma}$ of the induced Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta_e^*, \beta_\ell^*)$ that satisfies the following:*

(i) For each $(\theta, T_e) \in \Theta \times \mathcal{T}_e$, $u_e(\theta, \delta_e^*(\theta, T_e)) = \Pi_e(\hat{\sigma} | \theta, T_e, \beta_e^*)$.

[11] Recall that the space $\Theta \times Y \times M$ is finite.

(ii) For each $T_\ell \in \mathcal{T}_\ell$, $u_\ell(\delta^*_\ell(T_\ell)) = \Pi_\ell(\hat{\sigma}|T_\ell, \beta^*_\ell)$.

Proposition 4.1 provides a necessary and sufficient condition for (u_e, u_ℓ) to be a reduced form. In particular, it allows the analyst to restrict attention to Bayesian games that are induced by credible direct mechanisms and honest beliefs thereof. In fact, to verify that a set of reduced forms is a reduced form representation, we need only look at a subset of these mechanisms: mechanisms that also satisfy incentive compatibility and individual rationality conditions.

To formalize this, it will be useful to introduce some notation. Fix a reduced form (u_e, u_ℓ) and a credible extended direct mechanism $\mathcal{M}^d$. The expert's expected value of participating in $\mathcal{M}^d$ is

$$\mathcal{V}_e(\theta, \theta'|\mathcal{M}^d) := \sum_{(y_e, h_e) \in Y_e \times M_e} (u_e(\theta, h_e) + y_e) \cdot \text{marg}_{\,Y_e \times M_e} m(\theta')(y_e, h_e),$$

when beliefs are honest, the state is θ, and the report is θ'. Likewise, the layman's expected value of participating in $\mathcal{M}^d$ is

$$\mathcal{V}_\ell(\mathcal{M}^d) := \sum_{(y_\ell, h_\ell) \in Y_\ell \times M_\ell} (u_\ell(h_\ell) + y_\ell) \cdot \text{marg}_{\,Y_\ell \times M_\ell} \phi(y_\ell, h_\ell),$$

when beliefs are honest.

Definition 4.8. *A credible direct mechanism $\mathcal{M}^d$ is **Bayesian incentive compatible (BIC)** if, for each $\theta, \theta' \in \Theta$, $\mathcal{V}_e(\theta, \theta|\mathcal{M}^d) \geq \mathcal{V}_e(\theta, \theta'|\mathcal{M}^d)$.*

Definition 4.9. *A credible direct mechanism $\mathcal{M}^d$ is **individually rational (IR)** if $\mathcal{V}_\ell(\mathcal{M}^d) \geq \underline{\pi}_\ell$ and, for each $\theta \in \Theta$, $\mathcal{V}_e(\theta, \theta|\mathcal{M}^d) \geq \underline{\pi}_e(\theta)$.*

A credible direct mechanism $\mathcal{M}^d$ is BIC if the expert has incentives to report the state θ truthfully under honest beliefs. A credible direct mechanism $\mathcal{M}^d$ is IR if the agents' get a higher payoff by participating in $\mathcal{M}^d$ than by taking their outside option. The revelation principle in Rivera Mora [2021b] implies the following useful proposition.

Proposition 4.2. *A set of reduced forms* RF *is a reduced-form representation of G if and only if, for each mechanism $\mathcal{M}$ and each individually rational PBE (σ, β) of $(\mathcal{M}, G)$, there is a reduced form $(u_e, u_\ell) \in$ RF and a credible, BIC, IR direct mechanism $\mathcal{M}^d$ so that each honest PBE of $(\mathcal{M}^d, u_e, u_\ell)$ is equivalent to (σ, β).*

Proposition 4.1 states that to verify that (u_e, u_ℓ) is a reduced form, one must analyze only extended credible direct mechanisms. Proposition 4.2 provides a similar simplifying tool for reduced-form representations. Once the analyst knows that RF is a set of reduced forms, to verify RF is a reduced-form representation of G, it suffices to verify that equilibria across all supergames are captured by equilibria of only the psychological games associated with credible, BIC, and IR extended direct mechanisms.[12]

[12]So, Proposition 4.2 only requires looking at a subclass of credible extended direct mechanisms, while Proposition 4.1 requires looking at all credible extended direct mechanisms. To understand why the two requirements differ, note that the definition of a reduced form does not require that beliefs are generated by equilibrium behavior in the mechanism; however, the definition of a reduced-form representation does require such behavior. As a consequence, Proposition 4.2 can restrict attention to credible extended direct mechanisms satisfying BIC and IR, while Proposition 4.1 cannot.

5 Main Results

This section states the two main results. First, it provides a sufficient condition for a reduced form of G to allow for complete information sharing. Second, it provides a sufficient condition on a reduced-form representation of G to guarantee that no mechanism shares any relevant information.

5.1 Perfectly Revealing Games

The extent to which G is perfectly revealing—i.e., neutral mechanisms can induce the expert to completely reveal the state to the layman—depends on how a particular set of hierarchies impacts reduced forms of G. These are the hierarchies for which it is common belief that the layman is certain—but perhaps incorrect—about the state.

Fix $\theta \in \Theta$. Notice that there is a unique hierarchy profile $h^\theta = (h_e^\theta, h_\ell^\theta)$ so that $\eta_e(h_e^\theta)(h_\ell^\theta) = 1$ and $\eta_\ell(h_\ell^\theta)(\theta, h_e^\theta) = 1$. At the profile h^θ, there is common belief that "the layman believes that the state is θ." The **set of common degenerate hierarchies of beliefs** of i is $\mathrm{CDB}_i := \{h_i^\theta : \theta \in \Theta\}$. Write $h_i^\theta \geq h_i^{\theta'}$ if and only if $\theta \geq \theta'$.

Say that a game G is **perfectly revealing** if there is a mechanism $\mathcal{M}$ and an individually rational PBE (σ, β) of $(\mathcal{M}, G)$ so that, for each $(T_e, T_\ell) \in \mathcal{T}_e \times \mathcal{T}_\ell$ with $T_e \cap T_\ell \neq \emptyset$, there is $\theta \in \Theta$ with $\delta_e(\theta, T_e) = h_e^\theta$ and $\delta_\ell(T_\ell) = h_\ell^\theta$. That is, G is perfectly revealing if it is feasible to construct a mechanism and an individually rational PBE thereof, where, at each terminal information set, it is common knowledge that the layman learns the true state.

Definition 5.1. *Say that u_e is **supermodular on common degenerate beliefs** if, for each $\theta, \theta' \in \Theta$ with $\theta \geq \theta'$ and each $h_e, h_e' \in \mathrm{CDB}_e$ with $h_e \geq h_e'$,*

$$u_e(\theta, h_e) - u_e(\theta, h_e') \geq u_e(\theta', h_e) - u_e(\theta', h_e').$$

So, u_e is supermodular on CDB_e if the expert has a weakly higher incentive to induce "higher" common degenerate hierarchies when the state is high versus when it is low. Notice, in the example of Section 2, u_f satisfies increasing differences on $\{\underline{\theta}, \overline{\theta}\} \times \{0, 1\}$ when $c \leq 1$. This implies that u_f is supermodular on common degenerate beliefs.

Theorem 5.1. *Fix a reduced-form (u_e, u_ℓ) of G. If u_e is supermodular on common degenerate beliefs, then G is perfectly revealing.*

Theorem 5.1 implies that complete information sharing is possible if the expert has higher incentives to induce higher hierarchies when the state is high versus when the state is low. Remarkably, the result holds independent of the values that u_e takes outside CDB_e. This follows from the fact that the designer is able to design mechanisms with no ambiguous reports; the layman observes the report and each report is associated with a unique state. So, according to the equilibrium beliefs, the layman has degenerate beliefs about the state, even if the expert deviates from truthful reporting.

Notice, Theorem 5.1 establishes a sufficient condition for perfect revelation. Section 8.3 expands on this result by establishing a necessary and sufficient condition for perfect revelation (provided that G has a reduced form representation). However, it is often easier to verify supermodularity on common degenerate beliefs than to verify the alternate condition.

5.2 Concealing Games

Consider the benchmark where the expert and the layman do not interact before playing G. This is associated with a Bayesian game, in which the expert observes the state and, subsequently, agents select actions from A_e and A_ℓ. (No transfers are sent or received and the layman does not observe the state.) This **silent Bayesian game** provides a canonical benchmark to specify equilibrium payoffs absent information-sharing.

In the silent Bayesian game, the expert's strategy is $\sigma_e^s : \Theta \to \Delta(A_e)$ and the layman's strategy is $\sigma_\ell^s \in \Delta(A_\ell)$. We assume that the silent Bayesian game has some Bayesian equilibrium $\sigma^s = (\sigma_e^s, \sigma_\ell^s)$. Call $\Pi_e^s : \Theta \to \mathbb{R}$ and $\Pi_\ell^s \in \mathbb{R}$ **silent payoffs** if there is some Bayesian equilibrium σ^s with payoffs given by Π_e^s and Π_ℓ^s.[13]

Call the game G **concealing** if, for each mechanism $\mathcal{M}$ and each individually rational PBE (σ, β) of the supergame $(\mathcal{M}, G)$, there are silent payoffs (Π_e^s, Π_ℓ^s) so that $U_e(\sigma|\theta, \emptyset, \beta_e) = \Pi_e^s(\theta) + \mathbb{E}Y_e(\theta \mid \mathcal{M}, \sigma)$ and $U_\ell(\sigma|\{\emptyset\}, \beta_\ell) = \Pi_\ell^s + \mathbb{E}Y_\ell(\mathcal{M}, \sigma)$. So, G is concealing if mechanisms can only change the agents payoffs by changing the agents' transfers.

Note, the definition is silent about whether, in a concealing game, the layman learns information about Θ. In certain concealing games, the layman may learn inactionable information about Θ. To see this, consider the example of Section 2. Suppose that the prior μ assigns probability $\frac{1}{4}$ to the safe state $\overline{\theta}$. Section 2 argued that the game is concealing if $c > 1$. Consider a neutral mechanism where the firm chooses a cheap talk message in $M = \{\underline{m}, \overline{m}\}$. Consider a strategy of the firm σ_f satisfying $\sigma_f(\overline{\theta})(\overline{m}) = \frac{3}{4}$ and $\sigma_f(\underline{\theta})(\overline{m}) = \frac{5}{12}$. Under this strategy, the regulator's posteriors (that $\theta = 1$) are $p(\overline{m}) = \frac{3}{8}$ and $p(\underline{m}) = \frac{1}{8}$. So, the regulator optimally chooses *ban*, after observing both $\overline{m}$ and $\underline{m}$. That is, there is a Bayesian equilibrium (σ_f, σ_r), where $\sigma_r(\overline{m}) = \sigma_r(\underline{m}) = ban$. While the regulator still bans, the regulator's posteriors are not equal to his priors. So, there is some information shared.

A **higher-order statistic** (or a **statistic** for short) is a measurable and bounded function $f : H \to \mathbb{R}$. The reduced-from approach uses statistics to summarize the relevant dimensions of hierarchies of beliefs into a one-dimensional object. The definition of submodularity depends on how these statistics impact the reduced forms of G. So, it will be important to understand the equilibrium distribution of these statistics.

To describe the distribution of statistics, fix a credible direct mechanism $\mathcal{M}^d$ and let $\phi \in \Delta(\Theta \times Y \times M)$ be the ex-ante probability measure it induces. Write $\mathcal{F}$ for the discrete sigma-algebra on $\Theta \times Y \times M$. Note that each mapping $\mathbf{X} : \Theta \times Y \times M \to \mathbb{R}$ is a random variable on the probability space $(\Theta \times Y \times M, \mathcal{F}, \phi)$.[14] All random variables are denoted by bold capital letters.

[13]Notice that silent payoffs may be suitable outside options in some settings.

[14]Recall that $\Theta \times Y \times M$ is finite. So, all mappings $\mathbf{X} : \Theta \times Y \times M \to \mathbb{R}$ are measurable and have finite moments.

Write

$$\mathbb{E}_\phi[\mathbf{X}] = \int_{\Theta \times Y \times M} \mathbf{X}(\theta, y, h)\ d\phi,$$

for the expectation of $\mathbf{X}$. Similarly, write $\mathrm{Var}_\phi[\mathbf{X}]$ for the variance of $\mathbf{X}$ and $\mathrm{Cov}_\phi[\mathbf{X}, \mathbf{X}']$ for the the covariance of $\mathbf{X}$ and $\mathbf{X}'$. An important random variable will be $\boldsymbol{\Theta} = \mathrm{proj}_\Theta : (\Theta \times Y \times M) \to \Theta$, i.e., the random variable induced by the projection onto the state.

Fix a statistic $f : H \to \mathbb{R}$. Call $\mathbf{F} : (\Theta \times Y \times M) \to \mathbb{R}$ the **random variable associated with** f if $\mathbf{F}(\theta, y, h) = f(h)$ for each $(\theta, y, h) \in \Theta \times Y \times M$. So, $\mathbf{F}$ first restricts the statistic f to $M \subseteq H$ and then extends it to the space $\Theta \times Y \times M$.

Definition 5.2. *Let $f : H \to \mathbb{R}$ be a statistic with an associated random variable $\mathbf{F}$. Say f is* ***acute*** *if, for each credible direct mechanism $\mathcal{M}^d$ and induced ex-ante probability measure ϕ,*

(i) $\mathrm{Cov}_\phi[\mathbf{F}, \boldsymbol{\Theta}] \geq 0$, *and*

(ii) $\mathrm{Cov}_\phi[\mathbf{F}, \boldsymbol{\Theta}] = 0$ *implies $\mathbf{F}$ is ϕ-almost surely constant and equal to its prior value $f(\tilde{h})$.*[15]

Intuitively, a statistic f is acute if high values of f signal high values of θ. Importantly, acute statistics have a geometric interpretation. Write L^2 for the (quotient) normed space associated with $(\Theta \times Y \times M, \mathcal{F}, \phi)$, where each two random variables $\mathbf{X}$ and $\mathbf{Y}$ are equivalent if $\mathbf{X} - \mathbf{Y} = c$ almost surely for some $c \in \mathbb{R}$. In this space, $\sqrt{\mathrm{Var}_\phi[\cdot]}$ is a norm and $\mathrm{Cov}_\phi[\cdot, \cdot]$ is an inner product. In each credible direct mechanism, the random variable associated with an acute statistic satisfies one of two properties: Either $\mathbf{F}$ is almost surely constant—and thus equivalent to the zero vector—or the angle between $\boldsymbol{\Theta}$ and $\mathbf{F}$ is "acute" in the sense it measures strictly less than 90 degrees. (See Figure 5.1.)

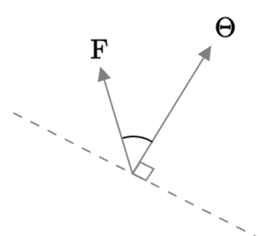


Figure 5.1. Angle between $\boldsymbol{\Theta}$ and $\mathbf{F}$ in the associated space L^2

We now provide a prominent example of an acute statistic.

Example 5.1. *Note that $\mathrm{marg}_\Theta \eta_\ell(h_\ell)$ is the layman's first-order beliefs about Θ given a hierarchy h_ℓ. Define $\mathbb{E}^1_\ell \theta : H_\ell \to \mathbb{R}$ by*

$$\mathbb{E}^1_\ell \theta(h_\ell) = \sum_{\theta \in \Theta} \theta \cdot \mathrm{marg}_\Theta \eta_\ell(h_\ell)(\theta).$$

The mapping $\mathbb{E}^1_\ell \theta$ is the layman's first-order expectation of the state given a hierarchy h_ℓ. Extend $\mathbb{E}^1_\ell \theta : H_\ell \to \mathbb{R}$ to $f^1 : H_e \times H_\ell \to \mathbb{R}$ by writing $f^1(h_e, h_\ell) = \mathbb{E}^1_\ell \theta(h_\ell)$. Lemma B.5 shows the statistic

[15] Recall $\tilde{h}$ is the hierarchy profile induced by the prior.

f^1 is acute. So, in each honest equilibrium of each credible direct mechanism, either the layman's expectation of the state is constant (i.e. the mechanism does not provide relevant information for the layman to update his conditional expectation) or positively correlated with the state (i.e. the mechanism does provide information for the layman to update his conditional expectation).

To understand why f^1 is acute, note that $\mathrm{Cov}_\phi[\mathbf{F}^1, \mathbf{\Theta}]$ captures an ex ante *relationship between the random variable that generates the state and the random variable that generates the conditional expectation. Intuitively,* ex ante*, a higher state should be associated with a higher conditional expectation of the state.*[16]

Lemmata B.5-B.8 show there are a plethora of acute statistics that are useful for applications. (In particular, they are used in the applications of Section 7.) These acute statistics correspond to increasing transformations of the layman's first-order expectation, the agents' higher-order expectations of the state, and positive linear combinations of all of them.

Acute statistics are useful because they impose Bayesian restrictions on how information flows from the expert to the layman: If agents Bayesian update, no credible direct mechanism can "deceive," in the sense of inducing lower values of the statistic for higher states. The next paragraph shows how each acute statistic induces an order over $\Delta(H_i)$. This order will be the key to define the submodularity condition.

Fix an acute statistic f. Given a hierarchy $h_i \in H_i$, write

$$\mathbb{E}_i f(h_i) = \int_{H_{-i}} f(h_i, h_{-i}) \; d\mathrm{marg}_{\,H_{-i}} \eta_i(h_i),$$

for i's expectation of f under h_i. Since f is bounded and measureable, $\mathbb{E}_i f$ is well defined, bounded, and measurable. Extend $\mathbb{E}_i f$ to a function $\mathbb{E}_i F : \Delta(H_i) \to \mathbb{R}$ so that

$$\mathbb{E}_i F(\nu) = \int_{H_i} \mathbb{E}_i f(h_i) \; d\nu.$$

So, $\mathbb{E}_i F(\nu)$ is i's expectation of f under a lottery of hierarchies $\nu \in \Delta(H_i)$. We can use f to define a complete order on $\Delta(H_e)$: Given $\nu, \nu' \in \Delta(H_e)$, write $\nu \geq_f \nu'$ (resp. $\nu >_f \nu'$) if $\mathbb{E}_e F(\nu) \geq \mathbb{E}_e F(\nu')$ (resp. $\mathbb{E}_e F(\nu) > \mathbb{E}_e F(\nu')$). So, $\nu \geq_f \nu'$ if the expert has a higher ex-ante expectation of the statistic f under ν than under ν'.[17]

Write

$$\mathbb{E} u_e(\theta, \nu) = \int_{H_e} u_e(\theta, h_e) \; d\nu,$$

for the expected value of $u_e : \Theta \times H_e \to \mathbb{R}$ given a state θ and a lottery of hierarchies $\nu \in \Delta(H_e)$.

Definition 5.3. *Fix an acute statistic f and $u_e : \Theta \times H_e \to \mathbb{R}$. Say $\mathbb{E}u_e$ is* ***strictly submodular*** *with respect to f if, for each $\theta, \theta' \in \Theta$ with $\theta > \theta'$ and each $\nu, \nu' \in \Delta(H_e)$ with $\nu >_f \nu'$,*

$$\mathbb{E}u_e(\theta, \nu) - \mathbb{E}u_e(\theta, \nu') < \mathbb{E}u_e(\theta', \nu) - \mathbb{E}u_e(\theta', \nu').$$

[16] Mathematically, the result follows from the fact that $\mathbf{F}^1$—the associated random variable to f^1—is a version of the conditional expectation of $\mathbf{\Theta}$. Thus, $\mathbf{F}^1$ is a projection of $\mathbf{\Theta}$ onto some subspace V of L^2. (See Theorem 4.1.15 in [Durrett, 2019].) Therefore, either $\mathbf{F}^1$ is a constant or the angle between $\mathbf{\Theta}$ and $\mathbf{F}^1$ is acute.

[17] It is obvious that different statistics can define different orders. However, it is also the case that different *acute* statistics can define different orders.

Strict submodularity with respect to f describes an expert who has strictly higher incentives to induce (in expectation) high values of f when the state θ is low. Since high values of f signal high values of θ, in a certain sense, the expert has strong preferences to "deceive the layman."

We now describe the final ingredient of the negative result. Fix a direct mechanism $\mathcal{M}^d$ with associated ex-ante probability measure ϕ. Say $\mathcal{M}^d$ is **not informative** about the statistic f if its associated random variable $\mathbf{F} = f(\tilde{h})$ ϕ-almost surely. So, $\mathcal{M}^d$ is informative if $\{(\theta, y, h) : \mathbf{F}(\theta, y, h) \neq f(\tilde{h})\}$ contains a set of strictly positive ϕ-measure. Say f is **essential** for a reduced form (u_e, u_ℓ) if, for each credible direct mechanism $\mathcal{M}^d$ that is not informative about f, there are silent payoffs (Π_e^s, Π_ℓ^s) so that $\mathcal{U}_e(\rho \mid \theta, \emptyset, \beta) = \Pi_e^s(\theta) + \mathbb{E}Y_e(\theta \mid \mathcal{M}^d, \sigma)$ and $\mathcal{U}_\ell(\rho \mid \{\emptyset\}, \beta) = \Pi_\ell^s + \mathbb{E}Y_\ell(\mathcal{M}^d, \sigma)$. That is, f is essential for (u_e, u_ℓ) if for each credible direct mechanism $\mathcal{M}^d$ either (1) $\mathcal{M}^d$ is informative about f or (2) agents get the their silent payoffs plus their expected transfers. So, the only way to change the agent's payoffs is by changing the value of an essential statistic f. An essential statistic f captures the information that is essential for the agents' payoffs.

Theorem 5.2. *Fix a reduced-form representation* RF *of G. If, for each $(u_e, u_\ell) \in$ RF, there is an acute statistic f so that f is essential for (u_e, u_ℓ) and $\mathbb{E}u_\ell$ is strictly submodular with respect to f, then G is concealing.*

Theorem 5.2 provides a sufficient condition to verify that G is concealing. If $\mathbb{E}u_e$ is strictly submodular with respect to an acute statistic f, then no credible direct mechanism is informative about f. Hence, since f is essential for (u_e, u_ℓ), the agents get their silent payoffs plus their expected transfers from the mechanism.

Notice how the theorem applies to the example of Section 2. In that example, the mapping ap$(\cdot)$ serves as a statistic that is essential for (u_f, u_r).[18] (If a direct mechanism $\mathcal{M}^d$ is not informative about ap$(\cdot)$, the regulator chooses *ban* and agents get their silent payoffs.) Moreover, by Lemma B.6, the statistic associated with ap$(\cdot)$ is a acute. (This follows from the the two equilibrium conditions stated in the example: (1) the action *approve* is weakly more likely when the state is $\overline{\theta}$, and (2) if the state does not impact the likelihood of *approve*, then the uniformed action (*ban)* is chosen.) In addition, if $c > 1$, then the expected reduced form $\mathbb{E}u_f$ is strictly submodular with respect to the statistic associated to ap$(\cdot)$. (This follows from Equation (1) and the fact that u_f is linear in ap$(\cdot)$.) Hence, Theorem 5.2 applies and G is concealing when $c > 1$.

6 Proofs of Main Results

6.1 Proof of Theorem 5.1

Fix a reduced form (u_e, u_ℓ) so that u_e is supermodular on common degenerate beliefs. It is sufficient to construct a credible, BIC, and IR extended direct mechanism $\mathcal{M}^d$—with respect to the reduced form (u_e, u_ℓ)—where the layman learns the state at all terminal nodes.

[18]Note, formally, ap is extended, so that its domain is hierarchies.

We begin by constructing the transfers used in the mechanism. First, set

$$y_\ell := \underline{\pi}_\ell - \sum_{\theta\in\Theta} u_\ell(\theta, h_\ell^\theta) \cdot \mu(\theta),$$

for the transfer of the layman. To construct the expert's transfers, it will be useful to introduce notation. Write $g : \Theta^2 \to \mathbb{R}$ for the function given by $g(\theta, \theta') = u_e(\theta, h_e^{\theta'})$. Since u_e is supermodular on common degenerate beliefs, g has increasing differences. Thus, by Lemmata B.9 and B.10, there exists a function $z : \Theta \to \mathbb{R}$ such that, for each $\theta, \theta' \in \Theta$,

$$g(\theta, \theta) + z(\theta) \geq g(\theta, \theta') + z(\theta'). \tag{2}$$

Moreover, since Θ is finite, there is a sufficiently large $c \in \mathbb{R}$ such that, for each $\theta \in \Theta$,

$$c + g(\theta, \theta) + z(\theta) - \underline{\pi}_e(\theta) \geq 0. \tag{3}$$

We use c and z to construct the expert's transfers, by setting $y_e^\theta := z(\theta) + c$ for each $\theta \in \Theta$. The set of transfers is $Y = \{y_e^\theta : \theta \in \Theta\} \times \{y_\ell\}$.

We now construct the direct mechanism. Set $M = \text{CDB}_e \times \text{CDB}_\ell$ and let $m : \Theta \to \Delta(Y \times M)$ be a protocol such that, for each report $\theta \in \Theta$, $m(\theta)((y_e^\theta, y_\ell), h^\theta) = 1$. Let $\mathcal{M}^d = (\Theta, (Y_i, M_i : i \in \{e, \ell\}), m)$ and notice that $\mathcal{M}^d$ is credible by construction.

Fix $\theta, \theta' \in \Theta$ and note that Equation (2) implies

$$g(\theta, \theta) + y_e^\theta \geq g(\theta, \theta') + y_e^{\theta'}.$$

Since $\mathcal{V}_e(\theta, \theta' \mid \mathcal{M}^d) = g(\theta, \theta') + y_e^{\theta'}$, it follows that $\mathcal{V}_e(\theta, \theta \mid \mathcal{M}^d) \geq \mathcal{V}_e(\theta, \theta' \mid \mathcal{M}^d)$. So, $\mathcal{M}^d$ is BIC. Equation (3) implies

$$g(\theta, \theta) + y_e^\theta \geq \underline{\pi}_e(\theta).$$

Hence, $\mathcal{V}_e(\theta, \theta \mid \mathcal{M}^d) \geq \underline{\pi}_e(\theta)$ for each $\theta \in \Theta$. So, $\mathcal{M}^d$ is IR for the expert. Notice also that, by definition of y_ℓ, $\mathcal{V}_\ell(\mathcal{M}^d) \geq \underline{\pi}_\ell$. So, $\mathcal{M}^d$ is IR for the layman.

Thus, the honest profile (ρ^*, β^*) is an individually rational PBE of $(\mathcal{M}^d, (u_e, u_\ell))$. Moreover, for each $(T_e, T_\ell) \in \mathcal{T}_e \times \mathcal{T}_\ell$ with $T_e \cap T_\ell \neq \emptyset$, there is $\theta \in \Theta$ so that $\delta_e^*(\theta, T_e) = h_e^\theta$ and $\delta_\ell^*(T_\ell) = h_\ell^\theta$. So, the layman learns the state at all terminal information sets.

6.2 Proof of Theorem 5.2

Fix an individually rational PBE (σ, β) of a supergame $(\mathcal{M}, G)$. Since RF is a reduced-form representation of G, there is a $(u_e, u_\ell) \in \text{RF}$, a credible, BIC, IR, extended direct mechanism $\mathcal{M}^d$ so that each honest assessment (ρ^*, β^*) of $(\mathcal{M}^d, u_e, u_\ell)$ is a PBE equivalent to (σ, β). (See Proposition 4.2.) Let f be an the acute statistic such that f is essential for (u_e, u_ℓ) and $\mathbb{E}u_\ell$ is strictly submodular with respect to f. So, it suffices to show that $\mathcal{M}^d$ is not informative about f. If so, then the agents get their silent payoffs plus the mechanisms transfers and, as a consequence, G is concealing.

Let $(\Theta \times Y \times H, \mathcal{B}, \phi)$ be the probability space induced by $\mathcal{M}^d$ and $\mathbf{F}$ be the random variable associated with f. Let $g : \Theta \to \mathbb{R}$ be the function defined by $g(\theta) = \mathbb{E}_\phi[\mathbf{F} | \mathbf{\Theta} = \theta]$. It suffices to show that g is weakly decreasing: If so, then $\text{Cov}_\phi[\mathbf{F}, \mathbf{\Theta}] \leq 0$. (See Lemma B.11.) Hence, given that f is acute, $\mathbf{F} = f(\tilde{h})$ ϕ-almost surely. Thus, $\mathcal{M}^d$ is not informative about f.

To show that g is weakly decreasing, define $\nu(\theta') = \text{marg}_{M_e} m(\theta')$ for each $\theta' \in \Theta$. So, $\nu(\theta')$ is the lottery of expert's hierarchies of beliefs that the mechanism selects when the expert reports θ'. Thus, for each $\theta, \theta' \in \Theta$,

$$\mathcal{V}_e(\theta, \theta' \mid \mathcal{M}^d) = \mathbb{E}u_e(\theta, \nu(\theta')) + \sum_{y_e \in Y_E} y_e \, \text{marg}_{Y_e} m(\theta')(y_e), \tag{4}$$

where the equality follows from the fact that $\mathcal{M}^d$ is credible, and, as a consequence, the expert's hierarchies coincide with her beliefs. (See Lemma 4.1 in Rivera Mora [2021b]).

Fix $\theta, \theta' \in \Theta$ with $\theta > \theta'$. Notice that BIC implies

$$\mathcal{V}_e(\theta, \theta \mid \mathcal{M}^d) - \mathcal{V}_e(\theta, \theta' \mid \mathcal{M}^d) \geq 0, \text{ and}$$
$$\mathcal{V}_e(\theta', \theta' \mid \mathcal{M}^d) - \mathcal{V}_e(\theta', \theta \mid \mathcal{M}^d) \geq 0.$$

Thus,

$$\mathcal{V}_e(\theta, \theta \mid \mathcal{M}^d) - \mathcal{V}_e(\theta, \theta' \mid \mathcal{M}^d) + \mathcal{V}_e(\theta', \theta' \mid \mathcal{M}^d) - \mathcal{V}_e(\theta', \theta \mid \mathcal{M}^d) \geq 0. \tag{5}$$

Substituting Equation (4) into Equation (5),

$$\mathbb{E}u_e(\theta, \nu(\theta)) - \mathbb{E}u_e(\theta, \nu(\theta')) + \mathbb{E}u_e(\theta', \nu(\theta')) - \mathbb{E}u_e(\theta', \nu(\theta)) \geq 0.$$

Since $\theta > \theta'$ and $\mathbb{E}u_e$ is strictly submodular, it follows that $\nu(\theta') \geq_f \nu(\theta)$, or equivalently, $\mathbb{E}_e F(\nu(\theta')) \geq \mathbb{E}_e F(\nu(\theta))$. Since $g(\theta) = \mathbb{E}_e F(\nu(\theta))$ and $g(\theta') = \mathbb{E}_e F(\nu(\theta'))$ (See Lemma B.12.) it follows that $g(\theta') \geq g(\theta)$, so g is decreasing.

7 Applications

This section provides examples of economically relevant games and characterizes when the games are perfectly revealing or concealing. In Applications 7.1 only the layman takes an action, and as a consequence, only first and second-order beliefs are relevant. In Application 7.2, both players take actions, and so, all hierarchies of beliefs are relevant.

7.1 Quadratic Payoffs with an Inactive Expert

The expert is a bureaucrat and the layman a politician. The bureaucrat has knowledge of the state of the world $\Theta \subsetneq \mathbb{R}$. The politician chooses a policy relevant to both agents. The bureaucrat and the politician have different policy preferences given the state of the world. This application characterizes when information sharing is feasible as a function of the type of disagreement in preferences about the right policy.

In this game $G = ((A_i, \pi_i) : i \in \{e, \ell\})$, the expert is inactive and both agents have a quadratic payoff structure.[19] So, the set of pure strategies for the expert is $A_e = \{\triangleleft\}$ (where $\triangleleft$ is a trivial

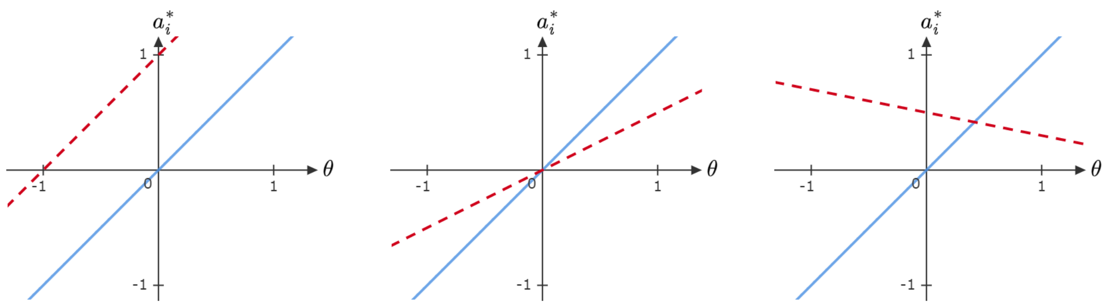
[19]Notice, while the expert is inactive in G, she can be active in the supergame.

action) and the set of pure strategies for the layman is $A_\ell = \mathbb{R}$. The payoffs functions are given by

$$\pi_\ell(\theta, a_\ell) = -(\theta - a_\ell)^2, \quad \text{and}$$
$$\pi_e(\theta, a_\ell) = -\lambda \cdot (b_1 + b_2\theta - a_\ell)^2,$$

where $\lambda \in \mathbb{R}_+$. The parameters $(b_1, b_2) \in \mathbb{R}^2$ describe the bias of the expert.

If the state were known, the layman's favorite action would be $a_\ell^*(\theta) = \theta$ and the expert's favorite action would be $a_e^*(\theta) = b_1 + b_2\theta$. So, $a_e^*(\cdot)$ is an affine transformation of $a_\ell^*(\cdot)$. Moreover, the agents would agree at all states if and only if $(b_1, b_2) = (0, 1)$. Figure 7.1 compares the agents' favorite actions under different forms of conflict. Panel (a) shows an example where the agents' level of conflict is constant in the state and Panel (b) an example where the level of conflict between the agents is increasing in the size of the state. Notice that in both cases the agents face "directional agreement." That is, they agree about the direction of how the action should change as the state changes. Panel (c) shows an example where the agents have "directional disagreement." That is, the agent's favorite actions move in opposite directions as the state changes.



(a) $b_1 = 1, b_2 = 0$ (b) $b_1 = 0, b_2 = \frac{1}{2}$ (c) $b_1 = \frac{1}{2}, b_2 = -\frac{1}{5}$

Figure 7.1. Agents' favorite action a_i^* for different values of the bias (b_1, b_2). The solid blue line represents the layman's favorite action. The dashed red line represent the expert's favorite action.

Proposition 7.1. *The game G has a reduced-form representation $RF = \{(u_e, u_\ell)\}$, where*

$$u_e(\theta, h_e) = -\lambda \int_{h_\ell \in H_\ell} \left(b_1 + b_2\theta - \mathbb{E}\theta_\ell^1(h_\ell)\right)^2 d\eta_e(h_e), \text{ and} \tag{6}$$

$$u_\ell(h_\ell) = -\sum_{\theta \in \Theta} \left(\theta - \mathbb{E}\theta_\ell^1(h_\ell)\right)^2 \operatorname{marg}_\Theta \eta_\ell(h_\ell)(\theta). \tag{7}$$

Moreover, the statistic $f^1 : H \to \mathbb{R}$ given by $f^1(h_e, h_\ell) = \mathbb{E}\theta_\ell^1(h_\ell)$ is acute and essential for (u_e, u_ℓ).

Recall that $\operatorname{marg}_\Theta \eta_\ell(\cdot) \in \Delta(\Theta)$ is the layman's first-order beliefs, $\mathbb{E}\theta_\ell^1(\cdot)$ is the layman's first-order expectation of the state, and $\eta_e(\cdot) \in \Delta(H_\ell)$ is the expert's beliefs about the layman's hierarchies of beliefs. Since payoffs are quadratic, the layman's unique equilibrium action corresponds to $\mathbb{E}\theta_\ell^1(\cdot)$. So, the belief-based utility associated with reduced form, u_ℓ, is the negative of his residual variance. Notice, the maximum value that u_ℓ can take is zero; that can only occur when the layman's first-order beliefs are degenerate. This reflects the layman's desire to learn the state. The expert cares about the action the layman will choose in G. Since the layman's equilibrium action depends on his first-order beliefs, the expert cares about her second-order ex-post beliefs. Her reduced-form utility is given by Equation (6); she would like the layman to believe that the state is close to her favorite action $a_e^*(\theta) = b_1 + b_2\theta$.

Proposition 7.2.

(i) If $b_2 \geq 0$ then the game G is perfectly revealing.

(ii) If $b_2 < 0$ then the game G is concealing.

Proof. We first show (i). Assume $b_2 \geq 0$. Let $\theta, \theta' \in \Theta$ and notice that $u_e(\theta, h_e^{\theta'}) = -\lambda(b_1+b_2\theta-\theta')^2$ has increasing differences with respect to θ and θ'. Thus, u_e is supermodular on common degenerate beliefs and the result follows from Theorem 5.1.

Now we show (ii). Notice that the statistic f^1 given by the layman's first-order expectation is acute. (See Lemma B.5.) Moreover, f^1 is essential for (u_e, u_ℓ). (See Proposition 7.1.) Thus, by Theorem 5.2, it suffices to show that $\mathbb{E}u_e$ is submodular with respect to f^1. First notice that

$$\begin{aligned}\mathbb{E}_e F^1(\nu) &= \int_{H_e} \mathbb{E}_e f^1(h_e)\ d\nu \\ &= \int_{H_e} \int_{H_\ell} f^1(h_e, h_\ell)\ d\eta_e(h_e)\ d\nu \\ &= \int_{H_e} \int_{H_\ell} \mathbb{E}_\ell(h_\ell)\ d\eta_e(h_e)\ d\nu. \end{aligned} \tag{8}$$

Thus, for each $\theta \in \Theta$ and $\nu \in \Delta(H_e)$,

$$\begin{aligned}\mathbb{E}u_e(\theta,\nu) &= -\lambda \int_{H_e} \int_{H_\ell} (b_1 + b_2\theta - \mathbb{E}_\ell(h_\ell))^2 d\eta_e(h_e)\ d\nu \\ &= -\lambda \int_{H_e} \int_{H_\ell} \left((b_1+b_2\theta)^2 + \mathbb{E}_\ell(h_\ell)^2 - 2\lambda(b_1+b_2\theta)\mathbb{E}_\ell(h_\ell)\right)\ d\eta_e(h_e)\ d\nu \\ &= -\lambda(b_1+b_2\theta)^2 - \int_{H_e} \int_{H_\ell} \lambda \mathbb{E}_\ell(h_\ell)^2\ d\eta_e(h_e)\ d\nu + 2\lambda(b_1+b_2\theta)\mathbb{E}_e F^1(\nu),\end{aligned}$$

where the last equality follows from Equation (8). Hence, if $\theta > \theta'$, $\mathbb{E}_e F^1(\nu) > \mathbb{E}_e F^1(\nu')$, and $b_2 < 0$, it follows that

$$\mathbb{E}u_e(\theta,\nu) - \mathbb{E}u_e(\theta',\nu) - \mathbb{E}u_e(\theta',\nu') + \mathbb{E}u_e(\theta,\nu') = \lambda b_2(\theta-\theta')(\mathbb{E}_e F^1(\nu) - \mathbb{E}_e F^1(\nu')) < 0.$$

So, $\mathbb{E}u_e$ is submodular with respect to f^1. □

Proposition 7.2 provides a complete taxonomy of the parameters that allows for or preclude information sharing. The extent to which information sharing is possible depends on the sign of b_2 and is independent of the parameter b_1. That is, the feasibility of information sharing does not depend on the absolute level of disagreement but rather on directional agreement.

Notice that Proposition 7.2 is silent about whether information sharing is desirable or not. Naturally, information sharing is always desirable by the layman since he is better off by making an informed decision. However, the expert may be worse off with information sharing depending on the bias. To analyze the trade-off between the agents' payoffs, consider the utilitarian welfare given by an exogenous information structure.[20] There are two important information structures: full information (i.e. the layman completely learns the state) and no information (i.e. the layman observes a non-informative signal.) Say that full revelation (resp. no information) is welfare maximizing it it maximizes the sum of the ex-ante agents' utilities among all information structures.

Proposition 7.3.

(i) If $b_2 \geq \frac{1}{2}$, then full information maximizes the payoffs of both agents.

(ii) If $b_2 \leq \frac{1}{2}$, then no information maximizes the expert's payoff, and full information maximizes the layman's payoffs.

(iii) If $\lambda(1 - 2b_2) \leq 1$, then full revelation maximizes welfare.

(iv) If $\lambda(1 - 2b_2) \geq 1$, then no information maximizes welfare.

Proposition 7.3 establishes the trade-off between the agent's preferences. Remarkably, the trade-off is irrelevant about the absolute level of disagreement (b_1). Moreover, the trade-off is characterized by the parameter b_2 and the expert's weight λ. In the region ($b_2 \geq \frac{1}{2}$), the expert agrees in how the layman reacts, (even though the expert may desire a different policy). So both agents get better off by revealing the state. In the region ($b_2 < \frac{1}{2}$), revealing information hurts the expert. This also holds when $b_2 \in [0, \frac{1}{2})$, even though the agents have directional agreement. In this case the expert has more moderate preferences than the layman. Thus, from the expert's perspective, layman overreacts to any information, resulting in the expert being hurt.

Propositions 7.2 and 7.3 imply that there is a conflict between what is feasible and what is optimal. The set of parameters is divided in four broad regions. (See Figure 7.2.) For positive b_2 and low λ, full revelation is feasible and optimal. There is directional agreement and the information helps the layman, the relatively important agent. For negative b_2 and high λ, information sharing is not feasible and not optimal. In this case information sharing hurts the expert, the relatively important agent. For negative b_2 and low λ, full information sharing is optimal but not feasible. Directional disagreement impedes information transmission which would benefit the layman, the relatively important agent. For positive but low b_2 and high λ, full information sharing is optimal but not feasible. In this case, the agents face directional agreement, but, from the expert's point of view, the layman overreacts to information. Hence a welfare-maximizer designer would prefer avoid information sharing even if he could implement it.

[20] An information structure is set of public signals S and an exogenous mapping $\chi : \Theta \to \Delta(S)$.

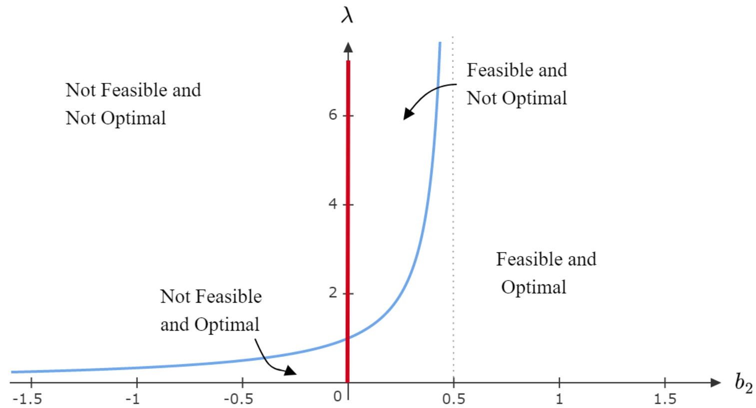


Figure 7.2. Feasibility and optimality of information sharing

7.2 Quadratic Payoffs with Active Agents

Two firms compete in a duopoly market. One of the firms (the expert) observes the state θ which captures the industry demand. (This could be because the firm purchased a forecast from a third party or for other reasons.) The second firm (the layman) does not observe θ. The application characterizes the extent to which the designer, (e.g. an industry association), can use neutral mechanisms so that the two firms share information.

The firms interaction is parametrized by a quadratic game G where both firms are active. Each firm chooses a real-valued action a_i. The payoff function of agent i is given by

$$\pi_i(\theta, a_i, a_{-i}) = \theta a_i - \tfrac{1}{2}a_i^2 + \alpha a_i a_{-i}, \tag{9}$$

where $\alpha \in (-1, 1)$ are commonly known parameters. The first term of (θa_i) represents i's benefit of the action a_i in terms of the demand level θ. The second term $(\frac{1}{2}a_i^2)$ represents the cost of increasing the action. The third term $(\alpha a_i a_{-i})$ represents the strategic interaction between the agents' actions.

Notice, when $\alpha < 0$, the game G captures a model of quantity competition with linear demand and constant marginal cost c. In that model, the action a_i is the quantity supplied by firm i. Firm i faces a linear inverse demand given by $P_i = (\theta - \frac{1}{2}a_i + \alpha a_{-i}) + c$. The profits of firm i are

$$(P_i - c)a_i = \theta a_i - \tfrac{1}{2}a_i^2 + \alpha a_i a_{-i},$$

as described by the function π_i. By contrast, when $\alpha > 0$, the game G captures a model of price competition with linear inverse demand and constant marginal cost c. In that model, the action a_i is firm i's the markup price, i.e., $a_i = p_i - c$. (Choosing a_i is equivalent of choosing p_i.) Firm i faces linear demand given by $Q_i = (\theta - \frac{1}{2}(p_i - c) + \alpha(p_{-i} - c)) = (\theta - \frac{1}{2}a_i + \alpha a_{-i})$. Thus, the profits of firm i are

$$(p_i - c)Q_i = \theta a_i - \tfrac{1}{2}a_i^2 + \alpha a_i a_{-i},$$

as described by π_i.[21]

Notice, if the state were known, i's favorite action (as a function of θ and a_{-i}) would be $a_i^*(\theta, a_{-i}) = \theta + \alpha a_{-i}$. So, if $\alpha < 0$ (quantity competition), then actions are strategic substitutes, i.e., the higher the action of the co-player, the greater the incentive to decrease one's own action. If $\alpha > 0$ (price competition), then actions are strategic complements, i.e., the higher the action of the co-player, the greater the incentive to increase one's own action.

In this game, both firms care both about the expectation of the state and the expectation of other the other firm's action. Hence, in contrast with the first application, firms care about their expectation of the state, their competitor's expectation of the state, their competitor's expectation of their expectation of the state, etc. So, to describe the associated psychological game, we need to introduce hierarchies of expectation.

Inductively define mappings $\mathbb{E}_i^k\theta : H_{-i} \to \mathbb{R}$ to be interpreted as firm i's k^{th}*-order expectation* (of the state). Specifically, assuming the mapping is defined for some k, define $\mathbb{E}_i^k\theta : H_{-i} \to \mathbb{R}$ so that

$$\mathbb{E}_{-i}^{k+1}\theta(h_{-i}) = \int_{H_i} \mathbb{E}_i^k(h_i) \cdot d\mathrm{marg}_{H_i}\eta_{-i}(h_{-i}).$$

Note, these mappings will depend on the hierarchies and, so, on the induced Bayesian game.

The key insight is that, in each induced Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$, the agents' behavior is characterized by their hierarchies of expectations. Intuitively, this follows by iteratively substituting the agents best response $a_i^* = \mathbb{E}_i[\theta + \alpha a_{-i}^*]$. In particular, the agent's payoffs can summarized in terms of a statistic $\overline{f} : H \to \mathbb{R}$ given by

$$\overline{f}(h_e, h_\ell) = (1+\alpha)\sum_{k=1}^{\infty} \alpha^{2k-2}\mathbb{E}_\ell^{2k-1}\theta(h_\ell).$$

This statistic is a weighted average of how the state is "commonly perceived" from the layman's perspective.[22] Lemma B.8 shows $\overline{f}$ is a convergent and a positive sum of acute statistics; as a consequence, it is acute.

Proposition 7.4. *The game G has a reduced-form representation* RF $= \{(u_e, u_\ell)\}$ *given by* $u_\ell(h_\ell) = \frac{1}{2}\left(\mathbb{E}_\ell\overline{f}(h_\ell)\right)^2$ *and* $u_e(\theta, h_e) = \frac{1}{2}\left(\theta + \alpha\mathbb{E}_e\overline{f}(h_e)\right)^2$*. Moreover, the statistic $\overline{f}$ is essential for (u_e, u_ℓ).*

Proposition 7.4 provides a simple way to characterize the firms' value of information. The firms only care about how information "moves" the statistic $\overline{f}$. Intuitively, this follows from the fact that i's optimal action is $a_i^* = \mathbb{E}_i[\theta + \alpha a_{-i}^*]$. So, in each induced Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$, the agents' unique equilibrium strategy is obtained by iteratively substituting their expected best responses. (Again, these expectations will be different, in different Bayesian games.) This iterated substitution of best responses is captured by $\overline{f}$. So, each induced Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ has a unique Bayesian equilibrium given by a (pure) strategy profile (σ_e, σ_ℓ) with $\sigma_e(\theta, T_e) = \theta + \alpha\mathbb{E}_e\overline{f}(\delta_e(\theta, T_e))$ and $\sigma_\ell(T_\ell) = \mathbb{E}_\ell\overline{f}(\delta_\ell(T_\ell))$. This leads to the reduced form of Proposition 7.4.

[21] Notice, the assumption $\alpha \in (-1, 1)$ states that the demand (resp. inverse inverse demand) for firm i is more sensible on its own action than in the other firm's action.

[22] Note, the statistic is a weighted sum of a first-, third-, fifth-... order belief of the layman.

The following proposition uses the reduced-form representation to characterize when G is perfectly revealing or concealing.

Proposition 7.5.

(i) If $\alpha \in [0,1)$ then the game G is perfectly revealing.

(ii) If $\alpha \in (-1,0)$, then the game G is concealing.

Proof. First we show (i). Assume $\alpha \in [0,1)$. Notice that if $h_e^\theta \in \mathrm{CDB}_e$ then,

$$\mathbb{E}_e \overline{f}(h_e^\theta)) = (1+\alpha)\sum_{k=1}^{\infty} \alpha^{2k-2}\theta = \tfrac{1}{1-\alpha}\theta,$$

So, for each $\theta, \theta' \in \Theta$, $u_e(\theta, h_e^{\theta'}) = \frac{1}{2}\left(\theta + \frac{\alpha}{1-\alpha}\theta'\right)^2$ satisfies weakly increasing differences with respect to θ and θ'. Hence, the result follows from Theorem 5.1.

We now show (ii). Assume $\alpha \in (-1,0)$. Notice that $\overline{f}$ is acute and essential for (u_e, u_ℓ). So, by Theorem 5.2, it suffices to show that $\mathbb{E}u_e$ is submodular with respect to $\overline{f}$. To show this, note that

$$\begin{aligned}
\mathbb{E}u_e(\theta,\nu) &= \tfrac{1}{2}\int_{H_e} \left(\theta + \alpha \mathbb{E}_e \overline{f}(h_e)\right)^2 \, d\nu \\
&= \tfrac{1}{2}\left(\theta^2 + 2\alpha\theta \int_{H_e} \mathbb{E}_\ell \overline{f}(h_\ell)\, d\nu + \alpha^2 \int_{H_e} (\mathbb{E}_e \overline{f}(h_e))^2 \, d\nu\right) \\
&= \tfrac{1}{2}\left(\theta^2 + 2\alpha\theta \mathbb{E}_\ell \overline{F}(\nu) + \alpha^2 \int_{H_e} (\mathbb{E}_e \overline{f}(h_e))^2 d\nu\right).
\end{aligned}$$

Thus, if $\theta > \theta'$ and $\mathbb{E}_e \overline{F}(\nu) > \mathbb{E}_e \overline{F}(\nu')$, then

$$\mathbb{E}u_e(\theta,\nu) - \mathbb{E}u_e(\theta',\nu) - \mathbb{E}u_e(\theta,\nu') + \mathbb{E}u_e(\theta',\nu') = \alpha(\theta-\theta')(\mathbb{E}_e\overline{F}(\nu) - \mathbb{E}_e\overline{F}(\nu')) < 0.$$

So, $\mathbb{E}u_e$ is submodular with respect to $\overline{f}$. □

Proposition 7.5 provides a complete taxonomy of the parameters that allows for or preclude information sharing. The feasibility of information sharing only depends on the sign of α, i.e., the strategic interaction of the firms' actions. So, the feasibility of information sharing only depends on the type of duopoly market the firms face. Under price competition, a "good-news firm" has a higher willingness to pay to induce optimistic beliefs (and to induce high market prices). Hence, there is a mechanism-contingent transfer scheme that induces the expert to reveal the state. By contrast, under quantity competition, the "good-news firm" has a higher willingness to pay to induce pessimistic beliefs (and to corner the market by inducing low quantities of the other firm). Thus, the submodularity condition is satisfied and information sharing is not possible.

Notice that Proposition 7.5 is silent about whether information sharing is desirable or not. Proposition B.2 in the appendix addresses its effect on firm profit. The layman firm strictly prefers full information revelation over no information being shared. But, information sharing can hurt the expert firm. Full information revelation increases (resp. decreases) the profits of the expert firm if $\alpha > 0$ (resp. $\alpha < 0$). Moreover, fully revealing the state increases (resp. decreases) total industry profits if $\alpha > 1-\sqrt{2}$ (resp. $\alpha < 1-\sqrt{2}$.). So, fully revealing the state increases industry profit (over no information sharing) only if there is price competition ($\alpha > 0$) or if there is quantity competition and the products are weak substitutes ($\alpha \in (1-\sqrt{2}, 0)$).

Remarkably, Propositions 7.5 and B.2 imply that there is a conflict between what is feasible and what is optimal. In the case of quantity competition with weak substitutes ($\alpha \in (1-\sqrt{2}, 0)$) full revelation of the state increases the industry profits, but it is not feasible under any neutral mechanism. This shows a limit on the use of neutral mechanisms. Increasing industry profit requires the use of non-neutral mechanisms.

8 Discussion

8.1 Existence of a Reduced-Form Representation

The main theorems assume the existence of reduced forms and reduced-form representations. In applications, it is easy to construct such belief-based utilities.

There are certain instances where a reduced form and a reduced-form representation are guaranteed. One is a case suggested in Section 4.1: if G is such that each associated Bayesian game has a unique equilibrium. More precisely, consider a set of finite type structures $\mathcal{S}$, $(S_e, S_\ell, b_e, b_\ell)$, where (i) each $S_i \subseteq \mathbb{R}$ is finite and (ii) $b_e : \Theta \times S_e \to \Delta(S_\ell)$ and $b_\ell : S_\ell \to \Delta(\Theta \times S_e)$ are belief maps. If, for each type structure in $\mathcal{S}$, the associated Bayesian game has a unique Bayesian equilibrium, then there is a reduced form. Moreover, if (u_e, u_ℓ) is a reduced form, then $\mathrm{RF} = \{(u_e, u_\ell)\}$ is a reduced-form representation. (See Lemma B.13.) A second case is when G involves an inactive expert (i.e., A_e is a singleton). Then the game has both a reduced form and a reduced-form representation, but the reduced-form representation need not be a singleton. (See Lemma B.14.) In each of the cases the respective lemmata show how to construct a reduced form representation.

However, there are games that have no reduced-form representation. The following example illustrates this.

Example 8.1. *Let Θ be a singleton, so there is no private information. The game G is Figure 4 in Aumann [1987]:*

	$\overline{a}_\ell$	$\underline{a}_\ell$
$\overline{a}_e$	$6, 6$	$2, 7$
$\underline{a}_e$	$7, 2$	$0, 0$

Figure 8.1. A game with no reduced form representation

There are three Nash equilibria, associated with payoff profiles $(2, 7)$, $(7, 2)$, and $(4\frac{2}{3}, 4\frac{2}{3})$. There is also a correlated equilibrium payoff profile of $(5, 5)$; it lies outside of the convex hull of Nash equilibria payoffs.

Proposition B.3 shows that the game has no reduced-form representation. To see this, note that Θ *and* $H = H_e \times H_\ell$ *are both singletons. So the set of reduced-forms can only capture Nash equilibrium payoffs. However, the set of correlated equilibrium payoffs of* G *is outside of the convex hull of the of Nash equilibrium payoffs. The key is that there is a neutral mechanism that can generate the correlated equilibrium payoff vector* $(5, 5)$*. This can be achieved by using a mechanisms that privately suggests actions that the agents should take. Hence, there are payoff profiles that can be generated by neutral mechanisms and cannot be captured by any reduced form.*

In Example 8.1, G may not have a reduced-form representation, if the hierarchies of beliefs are not sufficiently rich to capture all correlation that can be generated by external signals.[23] One approach is to extend the state space: from Θ to $\Theta \times \Phi$, where Φ represents payoff irrelevant information. I conjecture that, for finite games G, an extended state space generates both a reduced form and reduced-form representation.

8.2 Partial Information Sharing

Theorems 5.1 and 5.2 focus on the case where G is either super- or submodular. In the applications of Section 7, for each parametrization of the model, the game is either super- or submodular. So, for each parametrization, either the game is perfectly revealing or concealing.

There are games that are neither super- nor submodular. Thus, a game can be neither perfectly revealing nor concealing. To see this, consider the following example:

Example 8.2. *Let* $\Theta = \{1, 2, 3\}$*, with* $\mu(\theta) = \frac{1}{3}$ *for each* θ*. Consider a game where the expert is inactive and* $A_\ell = \mathbb{R}$*. Payoff functions are given by*

$$\pi_\ell(\theta, a_\ell) = -(\theta - a_\ell)^2$$

and

$$\pi_e(\theta, a_\ell) = -(\mathrm{b}(\theta) - a_\ell)^2,$$

where $\mathrm{b}(1) = 1, \mathrm{b}(2) = 3$ *and* $\mathrm{b}(3) = 2$*. Following Section 8.1, there is a reduced-form representation. But, Lemma B.16 shows that super- and submodularity do not hold.*

First observe that there is no mechanism that is perfectly revealing. The key is that, in any mechanism, an expert that observes the state $\theta = 2$ *(resp* $\theta = 3$*) tries to mimic an expert that observes the state* $\theta = 3$ *(resp.* $\theta = 2$*). However, there is a neutral mechanism that does reveal some information. Specifically, there is a mechanism that involves cheap talk messages, where the expert can credibly reveal if* $\theta = 1$ *or* $\theta \in \{2, 3\}$*. This information changes the layman's action and the payoffs of both agents. See Lemma B.16.*

8.3 Characterization of Perfectly Revealing Games

Theorem 5.1 provides a sufficient condition for complete information sharing. This section complements that result by providing necessary and sufficient conditions for complete information sharing,

[23] The fact that the hierarchies may not be sufficiently rich is reminiscent of the literature on redundant hierarchies. See Ely and Peski [2006] and Liu [2009].

provided that G has a reduced-form representation RF.

Following Vohra [2011], use Θ to define a completely connected network. Vertices in the network are given by Θ. Call a vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_{n+1}) \in \Theta^{n+1}$ a **path** (of size n). A path $\boldsymbol{\theta} = (\theta_1, ..., \theta_{n+1})$ is a **cycle** if $\theta_{n+1} = \theta_1$. Call a cycle $(\theta_1, ..., \theta_{n+1}) \in \Theta^{n+1}$ **simple** if $\theta_i \neq \theta_j$ for each $i < j$ with $i, j \in \{1, ..., n\}$, i.e, if the cycle visits each vertex at most once. A mapping $g : \Theta \times \Theta \to \mathbb{R}$ defines a flow cost of $g(\theta_i, \theta_i) - g(\theta_i, \theta_j)$ for the directed edge connecting θ_i with θ_j. Write

$$\mathcal{L}(g, (\theta_1, ..., \theta_{n+1})) = \sum_{k=1}^{n} [g(\theta_k, \theta_k) - g(\theta_k, \theta_{k+1})],$$

for the **length** of the cycle $(\theta_1, ..., \theta_{n+1})$ with respect of the function g. Say g satisfies **cyclical monotonicity** if $\mathcal{L}(g, \boldsymbol{\theta}) \geq 0$ for each cycle $\boldsymbol{\theta}$ of arbitrary size. One can verify that each cycle $\boldsymbol{\theta}$ can be "decomposed" into m simple cycles $\boldsymbol{\theta}^1, ..., \boldsymbol{\theta}^m$ with $\mathcal{L}(g, \boldsymbol{\theta}) = \sum_{k=1}^{m} \mathcal{L}(g, \boldsymbol{\theta}^k)$. So, the cyclical monotonicity condition can be verified by only analyzing the set of simple cycles.

Notice, cyclical monotonicity can be defined even if Θ is not an ordered set. If $\Theta \subseteq \mathbb{R}$ and g has weakly increasing differences then g satisfies cyclical monotonicity. (See Lemma B.9.) The following example shows that the converse does not hold.

Example 8.3. *Let* $\Theta = \{1, 2, 3\}$*. Let* $g : \Theta \times \Theta \to \mathbb{R}$ *be such that (1)* $g(2, 1) = g(2, 2) = 1$ *and (2)* $g(\theta, \theta') = 0$ *for* $(\theta, \theta') \in \Theta^2 \backslash \{(2, 1), (2, 2)\}$*. Note that* $C = \{(1, 2, 1), (1, 3, 1), (2, 3, 2), (1, 2, 3, 1), (1, 3, 2, 1)\}$ *is the set of all simply cycles for* Θ *(up to shifts). One can verify that* $\mathcal{L}(g, \boldsymbol{\theta}) \geq 0$ *for each cycle* $\boldsymbol{\theta} \in C$*. Hence,* g *satisfies cyclical monotonicity. However,* $g(2, 3) - g(2, 1) < g(1, 3) - g(1, 1)$*, so* g *does not satisfies weakly increasing differences.*

Fix a reduced form (u_e, u_ℓ). Say u_e **satisfies cyclical monotonicity on degenerate beliefs** if the function $g(\theta, \theta') = u_e(\theta, h_e^{\theta'})$ satisfies cyclical monotonicity. The following theorem characterizes when complete information-sharing is possible, given that G has a reduced-form representation RF.

Theorem 8.1. *Assume* G *has a reduced-form representation* RF*. The game* G *is perfectly revealing if and only if there is* $(u_e, u_\ell) \in \text{RF}$ *so that* u_e *satisfies cyclical monotonicity on degenerate beliefs.*

8.4 Psychological Motivations

The paper uses belief-based utility functions as an instrument to analyze information sharing in environments, even when agents have no intrinsic psychological motivations. However, the notions of supermodularity and submodularity can be used to characterize information sharing in situations when the belief-based utility functions model psychological motivations.

For instance, suppose a researcher (ℓ) seeks to elicit private information about a subject's (e) traits or characteristics. These traits can represent the subject's political ideology, religious beliefs, substance abuse, level of income, etc. Write $\Theta = \{0, 1\}$ for the set of traits and think of $\theta = 1$ as the trait that is perceived as "good."

The subject has image concerns—that is, he cares about whether the researcher perceives him as good ($\theta = 1$) or bad ($\theta = 0$). Write $p : H_\ell \to \mathbb{R}$ for the probability that a researcher with first-order belief $h_\ell^1 = \text{proj}_{\Delta(D_\ell^1)}(h_\ell)$ assigns to the good trait.[24] The subject's image concerns are modeled by a belief-based utility

$$u_e(\theta, h_e) = \int_{H_\ell} g(\theta, p(h_\ell))\ d\eta_e(h_e),$$

where $g : \Theta \times [0,1] \to \mathbb{R}$ is such that each $g(\theta, \cdot)$ is strictly increasing. So, independent of the subject's actual trait, the subject strictly prefers that the researcher believes $\theta = 1$. Importantly, the subject's image concerns may well depend on the subject's actual trait θ. (So, $g(0, \cdot) \neq g(1, \cdot)$.)

In this setting, non-neutral mechanisms are often out of reach. The only non-neutral mechanisms are mechanisms that depend on the subject's trait—something that, arguably, the designer does not know. So, it is natural to use neutral mechanisms to extract information about the subject's trait. Can neutral mechanisms induce the subject to reveal his private information?

Proposition 8.1.

(i) If $g : \Theta \times [0,1] \to \mathbb{R}$ satisfies weakly increasing differences on $\Theta \times \{0,1\}$, then there is a neutral mechanism where the researcher learns the true state θ.

(ii) If $g : \Theta \times [0,1] \to \mathbb{R}$ satisfies strictly decreasing differences on $\Theta \times [0,1]$, then in each neutral mechanism the layman's posterior is equal to the prior.

The proof uses the Revelation Principle in Rivera Mora [2021b] and Lemmata B.17 and B.18. The extent to which the researcher can vs. cannot learn the subject's private information depends on whether g has increasing or decreasing differences. If the "good" subject ($\theta = 1$) has (weakly) higher incentives to be perceived as "good," then there is a mechanism that results in complete information sharing. If the "bad" subject ($\theta = 0$) has strictly higher incentives to be perceived as "good," then no relevant information can be extracted.[25]

References

Charalambos D. Aliprantis and Kim C. Border. Springer, Berlin; London, 2006.

Robert J Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica: Journal of the Econometric Society*, pages 1–18, 1987.

Robert J Aumann and Sergiu Hart. Long cheap talk. *Econometrica*, 71(6):1619–1660, 2003.

David Austen-Smith and Jeffrey S Banks. Cheap talk and burned money. *Journal of Economic Theory*, 91(1):1–16, 2000.

Pierpaolo Battigalli and Martin Dufwenberg. Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35, 2009.

[24] Notice that p is measurable. To see this, notice that p is the composition of the projection of H_ℓ onto H_ℓ^1 and the mapping $\mu_\ell^1 \to \mu_\ell^1(1)$. Both mappings are measureable. (Apply Theorem 15.13 in Aliprantis and Border [2006].)

[25] Of course, there are functions g that differs from parts (i) and (ii). In that case, the researcher may be able to extract some but not complete information sharing.

Pierpaolo Battigalli, Emiliano Catonini, and Nicodemo De Vito. *Analysis of Strategic Thinking.* Typescript, Boconni University, 2022.

Elchanan Ben-Porath and Eddie Dekel. Signaling future actions and the potential for sacrifice. *Journal of Economic Theory*, 57(1):36–51, 1992.

Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

Andreas Blume, Oliver J Board, and Kohei Kawamura. Noisy talk. *Theoretical Economics*, 2(4): 395–440, 2007.

Kim Border. Comparing probability distributions. *Personal Notes*, 2001.

Adam Brandenburger and Eddie Dekel. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, 59(1):189–198, 1993.

Hans Carlsson and Eric Van Damme. Global games and equilibrium selection. *Econometrica: Journal of the Econometric Society*, pages 989–1018, 1993.

Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

Wouter Dessein. Authority and communication in organizations. *The Review of Economic Studies*, 69(4):811–838, 2002.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Piotr Dworczak. Mechanism design with aftermarkets: Cutoff mechanisms. *Econometrica*, 88(6): 2629–2661, 2020.

Jeffrey C Ely and Marcin Peski. Hierarchies of belief and interim rationalizability. *Theoretical Economics 1*, (1):19–65, 2006.

Esther Gal-Or. Information sharing in oligopoly. *Econometrica: Journal of the Econometric Society*, pages 329–343, 1985.

John Geanakoplos, David Pearce, and Ennio Stacchetti. Psychological games and sequential rationality. *Games and economic Behavior*, 1(1):60–79, 1989.

Stanton A Glantz, John Slade, Lisa A Bero, Peter Hanauer, and Deborah E Barnes. *The cigarette papers.* Univ of California Press, 1998.

Maria Goltsman, Johannes Hörner, Gregory Pavlov, and Francesco Squintani. Mediation, arbitration and negotiation. *Journal of Economic Theory*, 144(4):1397–1420, 2009.

Sanford J Grossman. The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, 24(3):461–483, 1981.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101 (6):2590–2615, 2011.

Navin Kartik. A note on cheap talk and burned money. *Journal of Economic Theory*, 136(1): 749–758, 2007.

Vijay Krishna and John Morgan. Contracting for information under imperfect commitment. *The RAND Journal of Economics*, 39(4):905–925, 2008.

Qingmin Liu. On redundant types and bayesian formulation of incomplete information. *Journal of Economic Theory*, 144(5):2115–2145, 2009.

Barry Meier. Origins of an epidemic: Purdue pharma knew its opiods were widely abused. *The New York Times*, 2018. URL https://www.nytimes.com/2018/05/29/health/purdue-opioids-oxycontin.html.

Jean-François Mertens and Shmuel Zamir. Formulation of bayesian analysis for games with incomplete information. *International journal of game theory*, 14(1):1–29, 1985.

Paul R Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, pages 380–391, 1981.

Stephen Morris. Political correctness. *Journal of political Economy*, 109(2):231–265, 2001.

Stephen Morris and Hyun Song Shin. Global games: Theory and applications. 2001.

Roger B Myerson. Optimal coordination mechanisms in generalized principal–agent problems. *Journal of mathematical economics*, 10(1):67–81, 1982.

Marco Ottaviani. The economics of advice. *University College London, mimeo*, 2000.

Marco Ottaviani and Peter Norman Sørensen. Reputational cheap talk. *The Rand journal of economics*, 37(1):155–175, 2006.

Michael Raith. A general model of information sharing in oligopoly. *Journal of economic theory*, 71(1):260–288, 1996.

Luis Rayo and Ilya Segal. Optimal information disclosure. *Journal of political Economy*, 118(5): 949–987, 2010.

Ernesto Rivera Mora. Information markets in games. *working paper*, 2021a.

Ernesto Rivera Mora. Mechanism design for belief-based preferences. *working paper*, 2021b.

Ariel Rubinstein. The electronic mail game: Strategic behavior under" almost common knowledge". *The American Economic Review*, pages 385–391, 1989.

Michael Spence. Job market signaling. In *Uncertainty in economics*, pages 281–306. Elsevier, 1978.

Ina Taneva. Information design. *American Economic Journal: Microeconomics*, 11(4):151–85, 2019.

Timothy Van Zandt and Xavier Vives. Monotone equilibria in bayesian games of strategic complementarities. *Journal of Economic Theory*, 134(1):339–360, 2007.

Xavier Vives. Duopoly information equilibrium: Cournot and bertrand. *Journal of economic theory*, 34(1):71–94, 1984.

Rakesh V Vohra. *Mechanism design: a linear programming approach*, volume 47. Cambridge University Press, 2011.

Georgia Wells, Jeff Horwitz, and Deepa Seetharaman. Facebook knows instagram is toxic for teen girls, company documents show. *The Wall Street Journal*, 2021. URL https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739.

Alexander Zimper. A note on the equivalence of rationalizability concepts in generalized nice games. *International Game Theory Review*, 8(04):669–674, 2006.

Amir Ziv. Information sharing in oligopoly: The truth-telling problem. *The RAND Journal of Economics*, pages 455–465, 1993.

Appendix A Hierarchy Mappings

Let W and X be two compact metric spaces endowed with their Borel sigma algebra and let $\rho : W \to X$ be a measurable mapping. Write $\underline{\rho} : \Delta(W) \to \Delta(X)$ for the function that maps each measure in $\Delta(W)$ to the image measure under $\rho : W \to X$. Notice that $\underline{\rho}$ is measurable. (See Theorem 15.14 in Aliprantis and Border [2006].)

Fix a mechanism $\mathcal{M}$ and interim belief functions $\beta_e : \Theta \times \mathcal{I}_e \to \Delta(V)$ and $\beta_\ell : \mathcal{I}_\ell \to \Delta(\Theta \times V)$. The belief mappings β_e, β_ℓ induce $\hat{\beta}_e : \Theta \times \mathcal{T}_e \to \Delta(\mathcal{T}_\ell)$ and $\hat{\beta}_\ell : \mathcal{T}_\ell \to \Delta(\Theta \times \mathcal{T}_e)$. (Recall that for each $(\theta, T_e, T_\ell) \in \Theta \times \mathcal{T}_e \times \mathcal{T}_\ell$, $\beta_e(\theta, T_e)(Z) = 1$ and $\beta_\ell(T_\ell)(Z) = 1$.)

Define mappings $\rho_e^1 : \mathcal{T}_\ell \to D_e^1$ and $\rho_\ell^1 : \Theta \times \mathcal{T}_e \to D_\ell^1$, so that $\rho_\ell^1(\theta, T_e) = \theta$. (Recall $D_e^1 = \{*\}$ so ρ_ℓ^1 is trivial.) Note that ρ_i^1 is measurable for each $i \in \{e, \ell\}$. Assume that the measurable maps ρ_e^k and ρ_ℓ^k are defined. Let $\rho_e^{k+1} : T_\ell \to D_e^{k+1}$ be defined so that $\rho_e^{k+1}(T_\ell) = (\rho_e^k(T_\ell), \underline{\rho_\ell^k}(\hat{\beta}_\ell(T_\ell)))$, Similarly, let $\rho_\ell^{k+1} : \Theta \times T_e \to D_\ell^{k+1}$ be defined so that $\rho_\ell^{k+1}(\Theta \times T_e) = (\rho_\ell^k(\Theta \times T_e), \underline{\rho}_e^k(\hat{\beta}_e(\Theta \times T_e)))$. Note that ρ_i^{k+1} is the composition of measurable functions and so, it is measurable.

Set $\delta_i^k := \underline{\rho}_i^k \circ \hat{\beta}_i$. Note, that for each T_ℓ, $\text{proj}_{D_e^k} \rho_e^{k+1}(T_\ell) = \rho_\ell^k(T_\ell)$. Thus, for each (θ, T_e), $\text{marg}_{D_e^k} \delta_e^{k+1}(\theta, T_e) = \delta_e^k(\theta, T_e)$. So, Write $\delta_e : \Theta \times T_e \to H_e$ for $\delta_e(\theta, T_e) = (\delta_e^1(\theta, T_e), \delta_e^2(\theta, T_e), ...)$. Similarly, for each T_ℓ, $\text{marg}_{D_\ell^k} \delta_\ell^{k+1}(T_\ell) = \delta_\ell^k(T_\ell)$. So, write $\delta_\ell : T_\ell \to H_\ell$ for $\delta_\ell(T_\ell) = (\delta_\ell^1(T_\ell), \delta_\ell^2(T_\ell), ...)$.

Appendix B Omitted Proofs

B.1 Proofs from Section 4

Fix a supergame $(\mathcal{M}, G)$, interim belief mappings β and belief-base utility functions (u_e, u_ℓ). A strategy profile (σ_e, σ_ℓ) of the supergame $(\mathcal{M}, G)$ induces the strategy profile $(\hat{\sigma}_e, \hat{\sigma}_\ell)$ of the Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ if $\hat{\sigma}_e$ (resp. $\hat{\sigma}_\ell$) is the restriction of σ_e (resp. $\hat{\sigma}_\ell$) to to $\Theta \times \mathcal{T}_e$ (resp. $\mathcal{T}_\ell$). Similarly, $\sigma = (\sigma_e, \sigma_\ell)$ induces the strategy profile (ρ_e, ρ_ℓ) of the psychological game $(\mathcal{M}, u_e, u_\ell)$ if ρ_e (resp. ρ_ℓ) is the restriction of σ_e (resp. σ_ℓ) to $\Theta \times (\mathcal{I}_e \backslash \mathcal{T}_e)$ (resp. $\mathcal{I}_\ell \backslash \mathcal{T}_\ell$).

Likewise, a strategy profile (ρ_e, ρ_ℓ) of the psychological game $(\mathcal{M}, u_e, u_\ell)$ and a strategy profile $(\hat{\sigma}_e, \hat{\sigma}_\ell)$ of the Bayesian game $BG(T_e, T_\ell, \beta)$ induce the strategy profile (σ_e, σ_ℓ) of the supergame $(\mathcal{M}, G)$ if

$$\sigma_e(\theta, I_e) = \begin{cases} \hat{\sigma}_e(\theta, I_e) & \text{if } I_e \in \mathcal{T}_e \\ \rho_e(\theta, I_e) & \text{if } I_e \in \mathcal{I}_e \backslash \mathcal{T}_e \end{cases} \quad \text{and} \quad \sigma_\ell(I_\ell) = \begin{cases} \hat{\sigma}_\ell(I_\ell) & \text{if } I_\ell \in \mathcal{T}_\ell \\ \rho_\ell(I_\ell) & \text{if } I_\ell \in \mathcal{I}_\ell \backslash \mathcal{T}_\ell. \end{cases}$$

Proof of Lemma 4.1. Fix a PBE (ρ, β) of the psychological game $(\mathcal{M}, u_e, u_\ell)$. Let $\delta_e : \Theta \times \mathcal{I}_e \to H_e$, $\delta_\ell : \mathcal{I}_\ell \to H_\ell$ be the hierarchy mappings associated with $\beta = (\beta_e, \beta_\ell)$. Let $\hat{\sigma} = (\hat{\sigma}_e, \hat{\sigma}_\ell)$ be the Bayesian Equilibrium associated to (u_e, u_ℓ) for the Bayesian game $BG(\mathcal{M}, \beta)$.

Write σ for the strategy profile of the supergame induced by the strategy profiles ρ and $\hat{\sigma}$. We first show payoff equivalence. Notice that for each $(\theta, I_e) \in \Theta \times \mathcal{I}_e$,

$$\begin{aligned} U_e(\sigma \mid \theta, I_e, \beta) &= \sum_{T_e \in \mathcal{T}_e} [\Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid \rho, \theta, I_e, \beta_e) \\ &= \sum_{T_e \in \mathcal{T}_e} [u_e(\theta, \delta_e(\theta, T_e)) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid \rho, \theta, I_e, \beta_e) \\ &= \mathcal{U}_e(\rho \mid \theta, I_e, \beta), \end{aligned}$$

where the second equality follows from the fact that $\hat{\sigma}$ is the BE associated to (u_e, u_ℓ). Moreover, $\mathbb{E}Y_e(\theta \mid \mathcal{M}, \sigma, \beta_e) = \mathbb{E}Y_\ell(\theta \mid \mathcal{M}, \rho, \beta_e)$. Using an analogous argument, for each $I_\ell \in \mathcal{I}_\ell$, $U_\ell(\sigma \mid I_\ell, \beta) = \mathcal{U}_\ell(\rho \mid I_\ell, \beta)$ and $\mathbb{E}Y_\ell(\mathcal{M}, \sigma, \beta_e) = \mathbb{E}Y_\ell(\mathcal{M}, \rho, \beta_e)$. So, (σ, β) and (ρ, β) are payoff equivalent.

It suffices to show that (ρ, β) is a perfect Bayesian equilibrium of $(\mathcal{M}, u_e, u_\ell)$. Notice that β is consistent with ρ, so β is also consistent with σ. Fix a strategy ρ'_e for the expert in the supergame and let σ'_e the supergame's strategy induced by $(\rho'_e, \hat{\sigma}_e)$. Then,

$$\begin{aligned}\mathcal{U}_e(\rho \mid \theta, I_e, \beta) &= U_e(\sigma \mid \theta, I_e, \beta) \\ &\geq U_e(\sigma'_e, \sigma_\ell \mid \theta, I_e, \beta), \\ &= \sum_{T_e \in \mathcal{T}_e} [\Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid (\rho'_e, \rho_\ell), \theta, I_e, \beta_e) \\ &= \sum_{T_e \in \mathcal{T}_e} [u_e(\theta, \delta_e(\theta, T_e)) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid (\rho'_e, \rho_\ell), \theta, I_e, \beta_e) \\ &= \mathcal{U}_e(\rho'_e, \rho_\ell \mid \theta, I_e, \beta),\end{aligned}$$

where the inequality follows from the fact that (σ, β) is a PBE of $(\mathcal{M}, G)$. An analogous argument establishes that for each strategy ρ'_ℓ and each information set $I_\ell \in \mathcal{I}_\ell$, $U_\ell(\sigma \mid I_\ell, \beta) \geq U_\ell(\sigma_e, \sigma'_\ell \mid I_\ell, \beta)$. So, sequential rationality is satisfied and (ρ, β) is a PBE of the $(\mathcal{M}, u_e, u_\ell)$. □

The proof of Proposition 4.1 requires an analysis in a particular class of direct mechanisms. Fix a direct mechanism $\mathcal{M}^d = ((M_i, Y_i : i \in \{e, \ell\}), m)$ with $Y_e = Y_\ell = \{0\}$. Write $\mathcal{T}_i^d$ for i's terminal information sets of $\mathcal{M}^d$. Write $T_e^d[\theta, h_e] = \{\theta\} \times \{h_e\} \times M_\ell \times Y$ for the expert's terminal information set associated with a report θ and hierarchy-message h_e. So, $\mathcal{T}_e^d = \{T_e^d[\theta, h_e] : (\theta, h_e) \in \Theta \times M_e\}$. Likewise, write $T_\ell^d[h_\ell] = \Theta \times M_e \times \{h_\ell\} \times Y$ for the layman's terminal information set associated with a hierarchy-message h_ℓ. So, $T_\ell^d = \{T_\ell^d[h_\ell] : h_\ell \in M_e\}$. Call the belief mappings $(\beta_e^*, \beta_\ell^*)$ **straight forward** for $\mathcal{M}^d$ if for each state $\theta \in \Theta$, report $\theta' \in \Theta$, and hierarchy messages $(h_e, h_\ell) \in M_e \times M_\ell$ the following is satisfied:

(i) $\beta_e^*(\theta, T_e^d[\theta', h_e])(T_\ell^d[h_\ell])) = \eta_e(h_e)(h_\ell)$.

(ii) $\beta_\ell^*(T_\ell^d[h_\ell])(\{\theta\} \times T_e^d[\theta, h_e])) = \eta_\ell(h_\ell)(\theta, h_\ell)$.

Lemma B.1. *Fix a direct mechanism $\mathcal{M}^d = ((M_i, Y_i : i \in \{e, \ell\}), m)$ with $Y_e = Y_\ell = \{0\}$, let $(\beta_e^*, \beta_\ell^*)$ be straight forward belief mappings for $\mathcal{M}^d$, and δ^* the associated hierachies. Then, for each $(\theta, \theta', h_e) \in \Theta \times \Theta \times M_e$, $\delta_e^*(\theta, T_e^d[\theta', h_e]) = h_e$ and for each $h_\ell \in M_\ell$, $\delta_\ell^*(T_\ell^d[h_\ell]) = h_\ell$. Moreover, if $\mathcal{M}^d$ is credible, then $\beta^* \in \mathrm{cons}(\mathcal{M}^d)$.*

Proof. First we show that for each $(\theta, \theta', h_e) \in \Theta \times \Theta \times M_e$, $\delta_e(\theta, T_e^d[\theta', h_e]) = h_e$ and for each $h_\ell \in M_\ell$, $\delta_\ell(\theta, T_e^d[\theta', h_e]) = h_e$. We show this by induction. Write $h_e = (\mu_e^1, \mu_e^2, ...)$ and $h_\ell = (\mu_\ell^1, \mu_\ell^2 ...)$ Notice by construction $\delta_e^1(\theta, T_e^d[\theta', h_e]) = \mu_e^1$ and for each $h_\ell \in M_\ell$, $\delta_\ell(\theta, T_e^d[\theta', h_e]) = \mu_\ell^1$. Moreover, if $\delta_e^k(\theta, T_e^d[\theta', h_e]) = \mu_e^k$ and for each $h_\ell \in M_\ell$, $\delta_\ell^k(\theta, T_e^d[\theta', h_e]) = \mu_\ell^k$, it follows that $\delta_e^{k+1}(\theta, T_e^d[\theta', h_e]) = \mu_e^{k+1}$ and for each $h_\ell \in M_\ell$, $\delta_\ell^{k+1}(\theta, T_e^d[\theta', h_e]) = \mu_\ell^{k+1}$, as desired.

Now we show the second part. Assume that $\mathcal{M}^d$ is credible. We now show that $\beta^* = (\beta_e^*, \beta_\ell^*)$ are interim honest beliefs consistent with the honest strategy profile ρ^* in the mechanism $\mathcal{M}^d$. Fix a state $\theta \in \Theta$, a report $\theta' \in \Theta$, and hierarchy messages $(h_e, h_\ell) \in M_e \times M_\ell$. To show consistency for the expert, it suffices to show that

$$\beta_e^*(\theta, T_e^d[\theta', h_e])(T_\ell^d[h_\ell])) \cdot \mathbb{P}[\{\theta\} \times T_e^d[\theta', h_e]) | \theta', \rho_\ell^*] = \mathbb{P}[\theta, h_e, h_\ell, y | \theta', \rho_\ell^*]. \tag{10}$$

To show this, note that given that the state is θ and the expert reports θ', the probability that the mechanism selects hierarchy-message h_e is

$$\mathbb{P}[\{\theta\} \times T_e^d[\theta', h_e])|\theta', \rho_\ell^*] = \operatorname{marg}_{M_e} m(\theta')(h_e).$$

So, multiplying both sides by $\mu(\theta')$ we get

$$\mu(\theta') \cdot \mathbb{P}[\{\theta\} \times T_e^d[\theta', h_e])|\theta', \rho_\ell^*] = \operatorname{marg}_{\Theta \times M_e} \phi(\theta', h_e).$$

Likewise, given that the state is θ and the expert reports θ', the probability that the mechanism selects hierarchy-messages are h_e and h_ℓ is

$$\mathbb{P}[\theta, h_e, h_\ell, y|\theta', \rho_\ell^*] = \operatorname{marg}_{M} m(\theta')(h_e, h_\ell).$$

(Recall that $Y = \{y\}$ with $y = (0, 0)$). So, multiplying both sides by $\mu(\theta')$ we get

$$\mu(\theta') \cdot \mathbb{P}[\theta, h_e, h_\ell, y|\theta', \rho_\ell^*] = \operatorname{marg}_{\Theta \times M} \phi(\theta', h_e, h_\ell).$$

Hence, since $\beta_e^*(\theta, T_e^d[\theta', h_e])(T_\ell^d[h_\ell])) = \eta_e(h_e)(h_\ell)$, Equation (10) holds if and only if

$$\eta_e(h_e)(h_\ell) \cdot \operatorname{marg}_{\Theta \times M_e} \phi(\theta', h_e) = \operatorname{marg}_{\Theta \times M_\ell} \phi(\theta', h_e, h_\ell),$$

which is satisfied by credibility of $\mathcal{M}^d$.

Consistency for the layman follows from a similar argument. Fix a state $\theta \in \Theta$ and hierarchy messages $(h_e, h_\ell) \in \Theta \times M_e \times M_\ell$. It suffices to show

$$\beta_\ell^*(T_\ell^d[h_\ell])(\{\theta\} \times T_e^d[\theta, h_e])) \cdot \mathbb{P}[\Theta \times T_\ell^d[h_\ell])|\rho_e^*] = \mathbb{P}[\theta, h_e, h_\ell, y|\rho_e^*].$$

Since $\beta_\ell^*(T_\ell^d[h_\ell])(\{\theta\} \times T_e^d[\theta, h_e])) = \eta_\ell(h_\ell)(\theta, h_e)$, this holds if and only if

$$\eta_\ell(h_\ell)(\theta, h_e) \cdot \operatorname{marg}_{M_\ell} \phi(h_\ell) = \operatorname{marg}_{\Theta \times M_\ell} \phi(\theta, h_e, h_\ell),$$

which is satisfied by credibility of $\mathcal{M}^d$. □

The proof of Proposition 4.1 requires the construction of credible direct mechanisms that general supergames induce. Fix a mechanism $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$, $\beta \in \operatorname{conv}(\mathcal{M})$, and let ρ be strategy profile in $\mathcal{M}$ such that β is consistent with ρ. Following Rivera Mora [2021b] we construct a credible direct mechanism $\mathcal{M}^d$ that $\mathcal{M}$ and β induce. We first construct the transfers and hierarchy-messages. Write $Y_i = \gamma_i(\mathcal{T}_i)$ for each $i \in \{e, \ell\}$. Let $\delta = (\delta_e, \delta_\ell)$ the hierarchy mappings associated to β and set $M_e = \delta_e(\Theta \times T_e)$ and $M_\ell = \delta_\ell(T_\ell)$. Note that $M = M_e \times M_\ell$ is belief closed. We now construct the protocol. Write $\mathcal{T}_e[\theta, h_e] = \{T_e \in T_e : \delta_e(\theta, T_e) = h_e\}$ and $\mathcal{T}_\ell[h_\ell] = \{T_\ell \in T_\ell : \delta_\ell(T_\ell) = h_\ell\}$. So, write

$$Z[\theta, y_e, y_\ell, h_e, h_\ell] = \{z \in Z : \gamma_e(T_e[z]) = y_e, \gamma_\ell(T_\ell[z]) = y_\ell, T_e[z] \in \mathcal{T}_e[\theta, h_e], \text{ and } T_\ell[z] \in \mathcal{T}_\ell[h_\ell]\},$$

for the terminal nodes that lead to transfers (y_e, y_ℓ) and hierarchies of beliefs (h_e, h_ℓ) when the state is θ. Define the protocol $m : \Theta \to \Delta(Y \times M)$ by $m(\theta)(y, h) = \mathbb{P}[Z[\theta, y, h]|\theta, \rho]$. That is, $m(\theta)(y, h)$ is probability of selecting an end node associated with transfers y and hierarchies h conditional on a state θ and strategy profile ρ.

Call $\mathcal{M}^d = ((M_i, Y_i : i \in \{e, \ell\}), m)$ the **extended direct mechanism induced by** $\mathcal{M}$ and β. Theorem 5.1 in Rivera Mora [2021b] shows that $\mathcal{M}^d$ is credible. The proof of Proposition 4.1 follows directly from the following stronger result.

Proposition B.1. *Belief-based utilities* (u_e, u_ℓ) *are a reduced-form of* G *if and only if, for each credible direct mechanism* $\mathcal{M}^d = ((M_i, Y_i : i \in \{e, \ell\}), m)$ *with* $Y_e = Y_\ell = \{0\}$ *and straightforward beliefs* $\beta^* = (\beta_e^*, \beta_\ell^*)$ *thereof, there is a strategy profile* $\hat{\sigma}$ *of the induced Bayesian game* $BG(\mathcal{T}_e^d, \mathcal{T}_\ell^d, \beta_e^*, \beta_\ell^*)$ *that satisfies the following:*

(i) For each $(\theta, \theta', h_e) \in \Theta \times \Theta \times M_e$, $\Pi_e(\hat{\sigma}^*|\theta, T_e^d[\theta', h_e], \beta_e^*) = u_e(\theta, h_e)$.

(ii) For each $h_\ell \in M_\ell$, $\Pi_\ell(\hat{\sigma}|T_\ell[h_\ell], \beta_\ell^*) = u_\ell(h_\ell)$.

Proof. The *only if* part follows directly from the fact that each direct mechanism $\mathcal{M}^d$ is itself a mechanism and each straight-forward beliefs β^* satisfies $\beta^* \in \text{cons}(\mathcal{M})$. (See Lemma B.1.)

We now show the converse. Fix a mechanism $\mathcal{M} = (\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ with $\beta \in \text{cons}(\mathcal{M})$. To show that (u_e, u_ℓ) is a reduced form, it suffices to construct a Bayesian equilibrium $\hat{\sigma}$ of $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ that satisfies

(iii) For each $(\theta, T_e) \in \Theta \times \mathcal{T}_e$, $\Pi_e(\hat{\sigma}|\theta, T_e, \beta_e) = u_e(\theta, \delta_e(\theta, T_e))$.

(iv) For each $T_\ell \in \mathcal{T}_\ell$, $\Pi_\ell(\hat{\sigma}|T_\ell, \beta_\ell) = u_\ell(\delta_\ell(T_\ell))$.

Notice, the set of Bayesian equilibria of $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ depend on the terminal information sets $(\mathcal{T}_e, \mathcal{T}_\ell)$ but not in the transfers of each terminal information set (i.e., the equilibria does not depend the mappings γ_i or the sets Y_i). So, it is without loss of generality assume assume $Y_i = \{0\}$ for each $i \in \{e, \ell\}$.

The construction of $\hat{\sigma}$ involves two steps. The first constructs $\hat{\sigma}$ and the second shows that $\hat{\sigma}$ is a Bayesian equilibrium that satisfies (iii) and (iv).

Step 1. Let $\mathcal{M}^d$ be the direct mechanism that $\mathcal{M}$ and β induce and let β^* be the straight forward interim beliefs of $\mathcal{M}^d$. By assumption, there is a Bayesian equilibrium $\hat{\sigma}^*$ of $BG(\mathcal{T}_e^d, \mathcal{T}_\ell^d, \beta^*)$ that satisfies (i) and (ii). We use the $\hat{\sigma}^*$ to construct $\hat{\sigma}$. For each $(\theta, T_e) \in \Theta \times T_e$ set $\hat{\sigma}(\theta, T_e) = \hat{\sigma}^*(\theta, T_e^d[\theta, h_e])$ where $h_e = \delta_e(\theta, T_e)$. Likewise, for each $T_\ell \in \mathcal{T}_\ell$ set $\hat{\sigma}(T_\ell) = \hat{\sigma}^*(T_\ell^d[h_\ell])$ where $h_\ell = \delta_\ell(T_\ell)$.

Step 2. To show that $\hat{\sigma}$ is a Bayesian equilibrium that satisfies (iii) and (iv), we first show some properties that (β_e, β_ℓ) and $(\beta_e^*, \beta_\ell^*)$ satisfy. If $T_e \in \mathcal{T}_e[\theta, h_e]$, then for each $h_\ell \in M_\ell$,

$$\sum_{T_\ell \in \mathcal{T}_\ell[h_\ell]} \beta_e(\theta, T_e)(T_\ell) = \eta_e(h_e)(h_\ell) = \beta_e^*(\theta, T_e^d[\theta, h_e])(T_\ell^d[h_\ell]), \tag{11}$$

where the first equality follows from Lemma B.3 in Rivera Mora [2021b] (and the fact that $\delta_e(\theta, T_e) = h_e$), and the second by definition of β_e^*. Similarly, if $T_\ell \in \mathcal{T}_\ell[h_\ell]$, then

$$\sum_{T_e \in \mathcal{T}_e[\theta, h_e]} \beta_\ell(T_\ell)(\theta, T_e) = \eta_\ell(h_\ell)(\theta, h_e) = \beta_e^*(T_\ell^d[h_\ell])(\theta, T_e^d[\theta, h_e]). \tag{12}$$

where the first equality follows from Lemma B.3 in Rivera Mora [2021b] (and the fact that $\delta_\ell(T_\ell) = h_\ell$), and the second by definition of β_ℓ^*.

We show that Equations (11) and (12) imply that the agents' expected payoffs in $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ under $\hat{\sigma}$ are equivalent to the expected payoffs in $BG(\mathcal{T}_e^d, \mathcal{T}_\ell^d, \beta^*)$ under $\hat{\sigma}^*$. We start with the expert. Fix $(\theta, T_e) \in \Theta \times \mathcal{T}_e[\theta, h_e]$ and a measurable $B_\ell \subseteq A_\ell$. Notice, given that ℓ plays according to $\hat{\sigma}_\ell$, the probability that an expert (θ, T_e) assigns to ℓ choosing an action in B_ℓ is

$$\sum_{h_\ell \in M_\ell} \sum_{T_\ell \in \mathcal{T}_\ell[h_\ell]} \beta_e(\theta, T_e)(T_\ell) \cdot \hat{\sigma}_\ell(T_\ell)(B_\ell) = \sum_{h_\ell \in M_\ell} \beta_e^*(\theta, T_e^d[\theta, h_e])(T_\ell^d[h_\ell]) \cdot \hat{\sigma}_\ell^*(T_\ell^d[h_\ell])(B_\ell),$$

where the equality follows from Equation (11) and the definition of $\hat{\sigma}_\ell$. Hence,

$$\hat{\Pi}_e(\hat{\sigma}|\theta, T_e, \beta_e) = \hat{\Pi}_e(\hat{\sigma}^*|\theta, T_e^d[\theta, h_e], \beta_e^*) = u_e(\theta, h_e).$$

Moreover, since $\hat{\sigma}_e^*(\theta, T_\ell^d[\theta, h_e]) \in \Delta(A_e)$ is optimal for $(\theta, T_\ell^d[\theta, h_e])$ in the game $BG(\mathcal{T}_e^d, \mathcal{T}_\ell^d, \beta^*)$, $\hat{\sigma}_e(\theta, T_\ell)$ is also optimal for (θ, T_ℓ) in the original game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$.

Now proceed with the layman. Fix $T_\ell \in \mathcal{T}_\ell[h_\ell]$ and a measurable $B_\ell \subseteq A_\ell$. Notice, given that e plays according to $\hat{\sigma}_e$, the probability that a layman T_ℓ assigns to state θ and e choosing an action in B_e is

$$\sum_{h_e \in M_e} \sum_{T_e \in \mathcal{T}_e[\theta, h_e]} \beta_\ell(T_\ell)(\theta, T_e) \cdot \hat{\sigma}_e(T_e)(B_e) = \sum_{h_e \in M_e} \beta_\ell^*(T_\ell)(\theta, T_e^d[\theta, h_e]) \cdot \hat{\sigma}_e^*(\theta, T_e^d[\theta, h_e])(B_e),$$

where the equality follows from Equation (12) and the definition of $\hat{\sigma}_e$. Hence,

$$\Pi_\ell(\hat{\sigma}|T_\ell, \beta_\ell) = \Pi_\ell(\hat{\sigma}^*|T_\ell^d[h_\ell], \beta_\ell^*) = u_\ell(h_\ell).$$

Moreover, since $\hat{\sigma}_\ell^*(T_\ell^d[h_\ell]) \in \Delta(A_\ell)$ is optimal for $T_\ell^d[h_\ell])$ in the game $BG(\mathcal{T}_e^d, \mathcal{T}_\ell^d, \beta^*)$, $\hat{\sigma}_\ell(T_\ell)$ is also optimal for T_ℓ in original game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. Therefore, $\hat{\sigma}$ is a Bayesian equilibrium that satisfies conditions (iii) and (iv). □

***Proof of Proposition** 4.2.* The "only if" part is trivial. We will show the "if" part. Assume that RF is a reduced form representation of G. Then, for each mechanism $\mathcal{M}$ and each individually rational PBE (σ, β) of $(\mathcal{M}, G)$, there is a mechanism $\mathcal{M}'$, a reduced form $(u_e, u_\ell) \in$ RF, and a PBE (ρ, β') of $(\mathcal{M}', (u_e, u_\ell))$ that is equivalent to (σ, β). So, by the revelation principle in Rivera Mora [2021b], there is a direct mechanism $\mathcal{M}^d$ and a honest PBE (ρ^*, β^*) of $(\mathcal{M}^d, u_e, u_\ell)$ that is equivalent to the PBE (ρ, β') of $(\mathcal{M}^d, u_e, u_\ell)$. Hence the PBE (ρ^*, β^*) of $BG(\mathcal{M}^d, u_e, u_\ell)$ is also equivalent to the PBE (σ, β) of $(\mathcal{M}, G)$. (See Lemma 4.1.) □

B.2 Proofs from Section 5

The lemmata in this section show there are a plethora of acute statistics that are useful for applications. In particular they are used in the applications of Section 7. These acute statistics correspond to monotone transformations of the first-order expectation, higher-order expectations of the state, and positive linear combinations of those. Before stating these lemmata we introduce some notation.

Fix a credible direct mechanism $\mathcal{M}^d$ and its induced probability space $(\Theta \times Y \times M, \mathcal{F}, \phi)$. For each odd $k \geq 1$, write $f^k(h_e, h_\ell) = \mathbb{E}_\ell^k \theta(h_\ell)$ and for each even $k \geq 2$ write $f^k(h_e, h_\ell) = \mathbb{E}_e^k \theta(h_e)$. Write $\mathbf{F}^0 = \mathbf{\Theta}$ and for each $k \in \mathbb{N}$ let $\mathbf{F}^k$ the random variable induced by f^k. Write $\mathcal{F}_e$ for the sigma algebra generated by $\text{proj}_{\Theta \times Y_e \times M_e}$ (which maps the expert's observables) and $\mathcal{F}_\ell$ for the sigma algebra generated by $\text{proj}_{Y_\ell \times M_\ell}$ (which maps the laymans's observables).

Lemma B.2. *Fix a direct mechanism $\mathcal{M}^d$ and its induced probability space $(\Theta \times Y \times M, \mathcal{B}, \phi)$. If $\mathcal{M}^d$ is credible then the following properties hold:*

(i) For each odd $k \geq 1$, $\mathbf{F}^k$ is a version of the conditional expectation of $\mathbf{F}^{k-1}$ given $\mathcal{F}_\ell$.

(ii) For each even $k \geq 2$, $\mathbf{F}^k$ is a version of the conditional expectation of $\mathbf{F}^{k-1}$ given $\mathcal{F}_e$.

Proof. First we show (1). Fix an odd $k \geq 1$ and note that $\mathbf{F}^k$ is $\mathcal{F}_\ell$-measurable. Fix $(y_\ell, h_\ell) \in Y_\ell \times M_\ell$ and write $A = \Theta \times Y_e \times \{y_\ell\} \times M_e \times \{h_\ell\}$ and $\mathbf{1}_{\boldsymbol{A}} : \Theta \times Y \times M \to \mathbb{R}$ the indicator random variable of the set A. Since $\mathcal{F}_\ell$ is generated for sets of this form, it suffices to show that $\mathbb{E}_\phi[\mathbf{F}^k \mathbf{1}_{\boldsymbol{A}}] = \mathbb{E}_\phi[\mathbf{F}^{k-1} \mathbf{1}_{\boldsymbol{A}}]$. If $k = 1$, then

$$
\begin{aligned}
\mathbb{E}_\phi[\mathbf{F}^1 \mathbf{1}_{\boldsymbol{A}}] &= \mathbb{E}^1_\ell \theta(h_\ell) \cdot \operatorname{marg} \phi(y_\ell, h_\ell) \\
&= \sum_{\theta' \in \Theta} \theta' \cdot \operatorname{marg}_\Theta \eta_\ell(h_\ell)(\theta') \cdot \operatorname{marg} \phi(y_\ell, h_\ell) \\
&= \sum_{\theta' \in \Theta} \theta' \cdot \operatorname{marg} \phi(\theta', y_\ell, h_\ell) \\
&= \mathbb{E}_\phi[\boldsymbol{\Theta} \mathbf{1}_{\boldsymbol{A}}],
\end{aligned}
$$

where the third equality follows from the credibility condition. If $k \geq 3$, then

$$
\begin{aligned}
\mathbb{E}_\phi[\mathbf{F}^k \mathbf{1}_{\boldsymbol{A}}] &= \mathbb{E}^k_\ell \theta(h_\ell) \cdot \operatorname{marg}_{Y_\ell \times M_\ell} \phi(y_\ell, h_\ell) \\
&= \sum_{h_e \in M_e} \mathbb{E}^{k-1}_e \theta(h_e) \cdot \operatorname{marg}_{M_e} \eta_\ell(h_\ell)(h_e) \cdot \operatorname{marg}_{Y_\ell \times M_\ell} \phi(y_\ell, h_\ell) \\
&= \sum_{h_e \in M_e} \mathbb{E}^{k-1}_e \theta(h_e) \cdot \operatorname{marg}_{Y_\ell \times M_e \times M_\ell} \phi(y_\ell, h_e, h_\ell) \\
&= \mathbb{E}_\phi[\mathbf{F}^{k-1} \mathbf{1}_{\boldsymbol{A}}],
\end{aligned}
$$

where the third equality follows from the credibility condition.

Now we show (2). Fix $(\theta, y_e, h_e) \in \Theta \times Y_e \times M_e$ and write $A = \{\theta\} \times \{y_e\} \times Y_\ell \times \{h_e\} \times M_\ell$ and $\mathbf{1}_{\boldsymbol{A}}$ the indicator function of A. Then, for each $k \geq 2$. Since F_ℓ is generated for sets of this form, it suffices to show that $\mathbb{E}_\phi[\mathbf{F}^k \mathbf{1}_{\boldsymbol{A}}] = \mathbb{E}_\phi[\mathbf{F}^{k-1} \mathbf{1}_{\boldsymbol{A}}]$.

$$
\begin{aligned}
\mathbb{E}_\phi[\mathbf{F}^k \mathbf{1}_{\boldsymbol{A}}] &= \mathbb{E}^k_e \theta(h_e) \cdot \operatorname{marg}_{\Theta \times Y_e \times M_e} \phi(\theta, y_e, h_e) \\
&= \sum_{h_e \in M_e} \mathbb{E}^{k-1}_\ell \theta(h_\ell) \cdot \eta_e(h_e)(h_\ell) \cdot \operatorname{marg}_{\Theta \times Y_e \times M_e} \phi(\theta, y_e, h_e) \\
&= \sum_{h_e \in M_e} \mathbb{E}^{k-1}_\ell \theta(h_\ell) \cdot \operatorname{marg}_{\Theta \times Y_\ell \times M_e \times M_\ell} \phi(\theta, y_e, h_e, h_\ell) \\
&= \mathbb{E}_\phi[\mathbf{F}^{k-1} \mathbf{1}_{\boldsymbol{A}}],
\end{aligned}
$$

where the third equality follows from the credibility condition. □

Lemma B.3. *Fix a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with $\mathbf{X}$ and $\mathbf{Y}$ two random variables with finite second moments thereof. Let $\mathcal{F}' \subseteq \mathcal{F}$ be a sigma algebra and $\mathbb{E}[\cdot|\mathcal{F}']$ a version of conditional expectation with respect to $\mathcal{F}'$. If $\mathbf{X}$ is $\mathcal{F}'$-measurable, then*

$$\operatorname{Cov}[\mathbf{X}, \mathbf{Y}] = \operatorname{Cov}[\mathbf{X}, \mathbb{E}[\mathbf{Y}|\mathcal{F}']].$$

Proof. Notice that

$$\begin{aligned}\mathrm{Cov}[\mathbf{X},\mathbf{Y}] &= \mathbb{E}[\mathbf{XY}] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{Y}] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{XY}|\mathcal{F}']] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbb{E}[\mathbf{Y}|\mathcal{F}']] \\ &= \mathbb{E}[\mathbf{X}\mathbb{E}[\mathbf{Y}|\mathcal{F}']] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbb{E}[\mathbf{Y}|\mathcal{F}']] \\ &= \mathrm{Cov}[\mathbf{X},\mathbb{E}[\mathbf{Y}|\mathcal{F}']],\end{aligned}$$

where the second equality follows from the law of iterated expectations and the third from the fact that $\mathbf{X}$ is $\mathcal{F}'$-measurable. (See Theorem 4.1.14 in Durrett [2019].)

Lemma B.4. *If $k, m, n \geq 0$ are so that $k + m \in \{2n, 2n-1\}$, then*

$$\mathrm{Cov}_\phi[\mathbf{F}^m, \mathbf{F}^k] = \mathrm{Cov}[\mathbf{F}^n, \mathbf{F}^n]. \tag{13}$$

Proof. Fix $n \in \mathbb{N}$. Without loss of generality, we will show the result for $k \geq m$. Notice that for each $d = 0, ..., 2n$ there is a unique pair (k, m) so that $k \geq m$, $k+m \in \{2n, 2n-1\}$, and $k - m = d$. We show the result holds for all of such pairs (k, m) by using induction over d.

If $d = 0$, then $m = k = n$ so then Equation (13) holds trivially. Assume that the result holds for the pair (k, m) associated to $d < 2n$. If $m \neq k \mod 2$, then there is $i \in \{e, \ell\}$ so that $\mathbf{F}^{m+1} = \mathbb{E}[\mathbf{F}^m|\mathcal{F}_i]$ and $\mathbf{F}^k$ is $\mathcal{F}_i$-measurable. Thus,

$$\mathrm{Cov}_\phi[\mathbf{F}^{m+1}, \mathbf{F}^k] = \mathrm{Cov}_\phi[\mathbf{F}^m, \mathbf{F}^k] = \mathrm{Cov}_\phi[\mathbf{F}^n, \mathbf{F}^n],$$

where the first equality follows from Lemma B.3 and the second equality follows from the induction hypothesis. If $m = k \mod 2$, then there is $i \in \{e, \ell\}$ so that $\mathbf{F}^k = \mathbb{E}[\mathbf{F}^{k-1}|\mathcal{F}_i]$ and $\mathbf{F}^m$ is $\mathcal{F}_i$-measurable.

$$\mathrm{Cov}_\phi[\mathbf{F}^m, \mathbf{F}^{k-1}] = \mathrm{Cov}_\phi[\mathbf{F}^m, \mathbf{F}^k] = \mathrm{Cov}_\phi[\mathbf{F}^n, \mathbf{F}^n],$$

where the first equality follows from Lemma B.3 and the second equality follows from the induction hypothesis. So, the result holds for $d + 1$. □

Lemma B.5. *Fix $k \in \mathbb{N}$. Let $i = e$ if k is even, and $i = \ell$ if k is odd. If $f^k : H \to \mathbb{R}$ is given by $f^k(h) = \mathbb{E}\theta_i^k(h_i)$ then f^k is an acute statistic.*

Proof. Let $n = \lfloor \frac{k+1}{2} \rfloor$ and $m = 0$ then $\mathrm{Cov}_\phi[\mathbf{F}^k, \mathbf{\Theta}] = \mathrm{Var}[\mathbf{F}^n] \geq 0$. (See Lemma B.4.) Moreover, if $\mathrm{Cov}_\phi[\mathbf{F}^k, \mathbf{\Theta}] = 0$ this implies $\mathrm{Var}_\phi[\mathbf{F}^n] = 0$. So $\mathbf{F}^k = \mathbb{E}_\phi[\mathbf{\Theta}] = f^k(\tilde{h})$ almost surely. □

Lemma B.6. *Fix a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with a random variable $\mathbf{X}$ with finite second moments and expected value $\mathbb{E}[\mathbf{X}] = c$. If $g : \mathbb{R} \to \mathbb{R}$ is weakly increasing and $g(\mathbf{X})$ has finite second moments, then $\mathrm{Cov}[\mathbf{X}, g(\mathbf{X})] \geq 0$. Moreover, $\mathrm{Cov}[\mathbf{X}, g(\mathbf{X})] = 0$ implies $g(\mathbf{X}) = g(c)$ almost surely.*

Proof. Write $c = \mathbb{E}[\mathbf{X}]$ and notice that

$$\begin{aligned}\mathrm{Cov}[\mathbf{X}, g(\mathbf{X})] &= \mathbb{E}[\mathbf{X}g(\mathbf{X})] - c\mathbb{E}[g(\mathbf{X})] \\ &= \mathbb{E}[(\mathbf{X} - c)g(\mathbf{X})] \\ &= \mathbb{E}[(\mathbf{X} - c)g(\mathbf{X})] - \mathbb{E}[\mathbf{X} - c]g(c) \\ &= \mathbb{E}[(\mathbf{X} - c)(g(\mathbf{X}) - g(c))].\end{aligned}$$

We show $\mathbb{E}[(\mathbf{X} - c)(g(\mathbf{X}) - g(c))] \geq 0$. Fix $\omega \in \Omega$. Since g is weakly increasing, $\mathbf{X}(\omega) \geq c$ implies $g(\mathbf{X}(\omega)) \geq g(c)$ and $\mathbf{X}(\omega) \leq c$ implies $g(\mathbf{X}(\omega)) \leq g(c)$. Thus, $(\mathbf{X}(\omega) - c)(g(\mathbf{X}(\omega)) - g(c)) \geq 0$. Moreover $(\mathbf{X}(\omega) - c)(g(\mathbf{X}(\omega)) - g(c)) = 0$ implies $g(\mathbf{X}(\omega)) - g(c)) = 0$. Thus $\mathbb{E}[(\mathbf{X} - c)(g(\mathbf{X}) - g(c))] \geq 0$ and $\mathbb{E}[(\mathbf{X} - c)(g(\mathbf{X}) - g(c))] = 0$ implies $g(\mathbf{X}) = g(c)$ almost surely. □

Lemma B.7. *Let $f^1 : H \to \mathbb{R}$ given by $f^1(h) = \mathbb{E}^1_\ell \theta(h_\ell)$ and $g : \mathbb{R} \to \mathbb{R}$ a weakly increasing function. Then $g \circ f^1$ is an acute statistic.*

Proof. Write $\mathbf{F}^1(\theta, y, h) = \mathbb{E}^1_\ell \theta(h_\ell)$ and notice that $\mathbf{F}^1$ and $g(\mathbf{F}^1)$ are both $\mathcal{F}_\ell$-measurable. By Lemma B.2, $\mathbf{F}^1$ is a version of the conditional expectation of $\boldsymbol{\Theta}$. Hence,

$$\mathrm{Cov}_\phi[g(\mathbf{F}^1), \boldsymbol{\Theta}] = \mathrm{Cov}_\phi[g(\mathbf{F}^1), \mathbf{F}^1] \geq 0,$$

where the first equality follows from Lemma B.3 and the second from Lemma B.6. Moreover, if $\mathrm{Cov}_\phi[g(\mathbf{F}^1), \mathbf{F}^1] = 0$ then $g(\mathbf{F}^1) = g(f(\tilde{h}))$ almost surely. □

Lemma B.8. *If $f(h) = (1 + \alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} \cdot \mathbb{E}^{2k-1}_\ell \theta(h_\ell)$, then f is an acute statistic.*

Proof. Let $\mathbf{F}$ the random variable associated with f. Write $\mathbf{X}^n = (1 + \alpha) \sum_{k=1}^{n} \alpha^{2k-2} \mathbf{F}^{2k-1}$ for each $k \in \mathbb{N}$ and notice that $\mathbf{F} = \lim_{n\to\infty} \mathbf{X}^n$.

Write $K = \max\{|\theta| : \theta \in \Theta\}$ and $\overline{K} = (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} K$. We show that all random variables in $\{\mathbf{X}^n : n \in \mathbb{N}\}$ are bounded by $\overline{K} \in \mathbb{R}$. Fix $n \in \mathbb{N}$ and notice that $|\mathbf{F}^k| \leq K$ for each $k \in \mathbb{N}$. Thus,

$$\begin{aligned}|\mathbf{X}^n| &= |(1 + \alpha) \textstyle\sum_{k=1}^{n} \alpha^{2k-2} \mathbf{F}^{2k-1}| \\ &\leq (1 + \alpha) \textstyle\sum_{k=1}^{n} \alpha^{2k-2} |\mathbf{F}^{2k-1}| \\ &\leq (1 + \alpha) \textstyle\sum_{k=1}^{n} \alpha^{2k-2} K \\ &\leq \overline{K}.\end{aligned}$$

Since $|\boldsymbol{\Theta}|$ is bounded by K, it follows that for each $n \in \mathbb{N}$, $|\boldsymbol{\Theta}\mathbf{X}^n|$ is bounded by $K\overline{K}$. Thus, by the dominated convergence theorem, $\mathbb{E}[\mathbf{F}\boldsymbol{\Theta}] = \mathbb{E}[\lim_{k\to\infty} \mathbf{X}^k \boldsymbol{\Theta}] = \lim_{k\to\infty} \mathbb{E}[\mathbf{X}^k \boldsymbol{\Theta}]$ and $\mathbb{E}[\mathbf{F}] =$

$\mathbb{E}[\lim_{k\to\infty} \mathbf{X}^k] = \lim_{k\to\infty} \mathbb{E}[\mathbf{X}^k]$. Hence,

$$\begin{aligned}
\text{Cov}_\phi[\mathbf{F}, \mathbf{\Theta}] &= \mathbb{E}_\phi[\mathbf{F}\mathbf{\Theta}] - \mathbb{E}_\phi[\mathbf{F}]\mathbb{E}[\mathbf{\Theta}] \\
&= \lim_{n\to\infty} (\mathbb{E}_\phi[\mathbf{X}^n\mathbf{\Theta}] - \mathbb{E}_\phi[\mathbf{X}^n]\mathbb{E}[\mathbf{\Theta}]) \\
&= \lim_{n\to\infty} \text{Cov}_\phi[\mathbf{X}^n, \mathbf{\Theta}] \\
&= \lim_{n\to\infty} (1+\alpha) \textstyle\sum_{k=1}^n \alpha^{2k-2} \text{Cov}_\phi[\mathbf{F}^k, \mathbf{\Theta}] \\
&\geq 0,
\end{aligned}$$

where the last equality from the fact that f^k is acute for each $k \in \mathbb{N}$. (See Lemma B.5.) Moreover, if $\text{Cov}_\phi[\mathbf{F}, \mathbf{\Theta}] = 0$, it follows that $\text{Cov}_\phi[\mathbf{F}^k, \mathbf{\Theta}] = 0$. So for each $k \in \mathbb{N}$, $\mathbf{F}^k = f^k(\tilde{h})$ a.s. which implies $\mathbf{F} = f(\tilde{h})$ a.s. Hence f is acute. □

B.3 Proofs from Section 6

Recall that $g : \Theta \times \Theta \to \mathbb{R}$ is cyclically monotone if $\sum_{k=1}^n g(\theta_k, \theta_k) - g(\theta_k, \theta_{k+1}) \geq 0$. for each cycle $(\theta_1, ..., \theta_n)$. (See Section 8.3.)

Lemma B.9. *If $g : \Theta \times \Theta \to \mathbb{R}$ is a function with weakly increasing differences, then it is cyclically monotone.*

Proof. The proof is by induction over the length n. For $n = 1$ the statement holds since in any cycle (θ_1, θ_2), it follows that $\theta_1 = \theta_2$ so $g(\theta_1, \theta_1) = g(\theta_1, \theta_2)$. We will show the statement holds for $n > 1$ provided it holds for $n - 1$. Fix a cycle $(\theta_1, ..., \theta_{n+1})$ with $\theta_{n+1} = \theta_1$. Without loss, assume that $\theta_n = \max\{\theta_1, ..., \theta_{n+1}\}$. (Otherwise shift the indexes of the cycle.) Now, consider the cycle $(\theta_1, ..., \theta_{n-1}, \theta_{n+1})$ of length $n - 1$. By the induction hypothesis, we have that

$$\sum_{k=1}^{n-2} (g(\theta_k, \theta_k) - g(\theta_k, \theta_{k+1})) + g(\theta_{n-1}, \theta_{n-1}) - g(\theta_{n-1}, \theta_{n+1}) \geq 0.$$

In addition, weakly increasing differences and $\theta_n \geq \max\{\theta_{n-1}, \theta_{n+1}\}$, implies

$$g(\theta_n, \theta_n) + g(\theta_{n-1}, \theta_{n+1}) - g(\theta_n, \theta_{n+1}) - g(\theta_{n-1}, \theta_n) \geq 0.$$

So, adding the two equalities we get

$$\sum_{k=1}^{n} g(\theta_k, \theta_k) - g(\theta_k, \theta_{k+1}) \geq 0.$$

So, the statement holds for n. □

Lemma B.10. ***(Rochet)*** *Fix finite set $\Theta \subsetneq \mathbb{R}$ and function $g : \Theta \times \Theta \to \mathbb{R}$. The function g satisfies cyclical monotonicity if and only if there exist a function $z : \Theta \to \mathbb{R}$ such that $g(\theta, \theta) + z(\theta) \geq g(\theta, \theta') + z(\theta')$ for each $\theta, \theta' \in \Theta$.*

Proof. Let $\text{Id} : \Theta \to \Theta$ be the identity function. Then g satisfies cyclical monotonicity if and only if the graph induced by (g, Id) has no finite cycles of negative length. The result follows from Theorem 4.2.1 in Vohra [2011]. □

Lemma B.11. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be real random variables with finite second moments and let* $\mathbb{E}[\mathbf{X} \mid \mathbf{Y}]$ *be a version of the conditional expectation of* $\mathbf{X}$ *given the sigma-algebra generated by* $\mathbf{Y}$. *If the function* $g(y) = \mathbb{E}[\mathbf{X} \mid \mathbf{Y} = y]$ *is decreasing, then* $\mathrm{Cov}[\mathbf{X}, \mathbf{Y}] \leq 0$.

Proof. By the law of iterated expectations,

$$\mathbb{E}[\mathbf{XY}] = \mathbb{E}[\mathbb{E}[\mathbf{XY} \mid \mathbf{Y}]] = \mathbb{E}[\mathbf{Y}\mathbb{E}[\mathbf{X} \mid \mathbf{Y}]] = \mathbb{E}[\mathbf{Y}g(\mathbf{Y})],$$

and similarly $\mathbb{E}[\mathbf{X}] = \mathbb{E}[\mathbb{E}[\mathbf{X} \mid \mathbf{Y}]] = \mathbb{E}[g(\mathbf{Y})]$. Thus,

$$\begin{aligned}\mathrm{Cov}[\mathbf{X}, \mathbf{Y}] &= \mathbb{E}[\mathbf{XY}] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{Y}] \\ &= \mathbb{E}[\mathbf{Y}g(\mathbf{Y})] - \mathbb{E}[g(\mathbf{Y})]\ \mathbb{E}[\mathbf{Y}] \\ &= \mathrm{Cov}[\mathbf{Y}, g(\mathbf{Y})].\end{aligned}$$

So is sufficient to show $\mathrm{Cov}[\mathbf{Y}, g(\mathbf{Y})] \leq 0$. Now, let $\mathbf{Z}$ be an independent random variable identically distributed as $\mathbf{Y}$. Since g is non-increasing, for each $y, z \in \mathbb{R}$, $(y - z)(g(y) - g(z)) \leq 0$. So $\mathbb{E}[(\mathbf{Y} - \mathbf{Z})(g(\mathbf{Y}) - g(\mathbf{Z}))] \leq 0$. Then, by independence of $\mathbf{Y}$ and $\mathbf{Z}$,

$$\mathbb{E}[\mathbf{Y}g(\mathbf{Y})] + \mathbb{E}[\mathbf{Z}g(\mathbf{Z})] - \mathbb{E}[\mathbf{Y}]\mathbb{E}[g(\mathbf{Z})] - \mathbb{E}[\mathbf{Z}]\mathbb{E}[g(\mathbf{Y})] \leq 0.$$

Therefore, since $\mathbf{Y}$ is identically distributed as $\mathbf{Z}$, $2\mathbb{E}[\mathbf{Y}g(\mathbf{Y})] - 2\mathbb{E}[\mathbf{Y}]\mathbb{E}[g(\mathbf{Y})] \leq 0$. Therefore,

$$\mathrm{Cov}[\mathbf{Y}, g(\mathbf{Y})] = \mathbb{E}[\mathbf{Y}g(\mathbf{Y})] - \mathbb{E}[\mathbf{Y}]\mathbb{E}[g(\mathbf{Y})] \leq 0,$$

as desired. □

Lemma B.12. *Fix a statistic* $f : H \to \mathbb{R}$ *and a direct mechanism* $\mathcal{M}^d$ *with protocol* m *and associated ex-ante probability measure* ϕ. *If* $\nu : \Theta \to \Delta(H_i)$ *is given by* $\nu(\theta) = \mathrm{marg}_{H_e} m(\theta)(h_e)$ *then* $\mathbb{E}F_i(\nu(\theta)) = \mathbb{E}_\phi[\mathbf{F}|\mathbf{\Theta} = \theta]$.

Proof. Write $\mathbf{H} : \Theta \times Y \times M \to M$ for the projection of the probability space onto M. It suffices to show that $\mathbb{E}F_i(\nu(\theta)) = \mathbb{E}_\phi[f(\mathbf{H}) \mid \mathbf{\Theta} = \theta]$. Notice that for each $(\theta, h) \in \Theta \times M$

$$\mathrm{marg}_M m(\theta)(h) = \frac{\mathrm{marg}_{\Theta \times M}\phi(\theta, h)}{\mathrm{marg}_\Theta \phi(\theta)} = \mathbb{P}_\phi[\mathbf{H} = h \mid \mathbf{\Theta} = \theta], \tag{14}$$

where $\mathbb{P}_\phi[\mathbf{H} = h \mid \mathbf{\Theta} = \theta]$ denotes the conditional probability of $\mathbf{H}$ conditional on $\mathbf{\Theta} = \theta$. Then,

$$\begin{aligned}\mathbb{E}F_i(\nu(\theta)) &= \int_{H_i} \mathbb{E}F_i(h_i)\ d\nu(\theta) \\ &= \sum_{h_i \in M_i} \sum_{h_{-i} \in M_{-i}} f(h_i, h_{-i}) \cdot \mathrm{marg}_H m(\theta)(h_i, h_{-i}) \\ &= \sum_{(y,h) \in Y \times M} f(h) \cdot m(\theta)(y, h) \\ &= \sum_{h \in M} f(h) \cdot \mathbb{P}_\phi[\mathbf{H} = h \mid \mathbf{\Theta} = \theta] \\ &= \mathbb{E}_\phi[f(\mathbf{H}) \mid \mathbf{\Theta} = \theta].\end{aligned}$$

where the fourth equality follows from Equation (14). □

B.4 Proofs from Section 7

Proof of Proposition 7.1. The proof is divided in three steps. The first shows that (u_e, u_ℓ) is a reduced form of G. The second shows that RF $= \{(u_e, u_\ell)\}$ is a reduced form representation for G. The third shows that f^1 is an essential statistic for (u_e, u_ℓ).

Step 1. Fix a Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. Fix a strategy profile $\hat{\sigma}$. In this Bayesian, game, the expected payoff of the layman from choosing action $s_\ell \in \mathbb{R}$ is

$$\sum_{\theta\in\Theta} -(\theta - s_\ell)^2 \cdot \text{marg}_\Theta \beta_\ell(T_\ell)(\theta).$$

Thus, $\hat{\sigma}$ is a PBE if and only if $\hat{\sigma}_e$ is pure strategy with

$$\hat{\sigma}_\ell(T_\ell) = \sum_{\theta\in\Theta} \theta \cdot \text{marg}_\Theta \beta_\ell(T_\ell)(\theta) = \mathbb{E}^1_\ell \theta(\delta_\ell(T_\ell)). \tag{15}$$

Hence, for each $T_\ell \in \mathcal{T}_\ell$,

$$\begin{aligned}
\Pi_\ell(T_\ell|\hat{\sigma}, \beta) &= \sum_{\theta\in\Theta} -(\theta - \hat{\sigma}_\ell(T_\ell))^2 \cdot \text{marg}_\Theta \beta_\ell(T_\ell)(\theta) \\
&= \sum_{\theta\in\Theta} -(\theta - \mathbb{E}^1_\ell \theta(\delta_\ell(T_\ell)))^2 \cdot \text{marg}_\Theta \eta_\ell(\delta_\ell(T_\ell))(\theta) \\
&= u_\ell(\delta_\ell(T_\ell)).
\end{aligned}$$

Similarly, for each $(\theta, T_e) \in \Theta \times \mathcal{T}_e$,

$$\begin{aligned}
\Pi_e(\theta, T_e|\hat{\sigma}, \beta) &= \sum_{T_\ell\in\mathcal{T}_\ell} -w(\theta - b_1 - b_2\sigma_\ell(T_\ell))^2 \beta_e(\theta, T_e)(T_\ell) \\
&= \int_{H_\ell} -w(\theta - b_1 - b_2\mathbb{E}^1_\ell \theta(h_\ell))^2 \; d\eta_e(\delta_e(\theta, T_e)) \\
&= u_e(\theta, \delta_e(\theta, T_\ell)),
\end{aligned}$$

where the second equality follows from the definition of η_e and δ_e. Therefore, (u_e, u_ℓ) is a reduced-form of G.

Step 2. By step 1, each induced game BG$(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ has a unique equilibrium given by the strategy profile $\hat{\sigma}$ that satisfies Equation (15). Hence, RF $= \{(u_e, u_\ell)\}$ is a reduced-form representation. (See Lemma B.13.)

Step 3. We show that f^1 is essential for (u_e, u_ℓ). Fix a credible direct mechanism $\mathcal{M}^d$, let (σ^*, β^*) the honest profile and let ϕ the associated ex-ante probability measure it induces. If $\mathcal{M}^d$ is not informative about f, it follows that $\mathbf{F}^1 = f^1(\tilde{h})$ almost surely. As a consequence

$$\mathcal{U}_e(\rho^*|, \theta, \{\emptyset\}, \beta^*) = \Pi^s_e(\theta) + \mathbb{E}Y_e(\theta|\mathcal{M}^d, \sigma), \text{ and}$$
$$\mathcal{U}_\ell(\rho^*|\{\emptyset\}, \beta^*) = \Pi^s_\ell + \mathbb{E}Y_\ell(\theta|\mathcal{M}^d, \sigma),$$

where $\Pi^s_e(\theta) = -(\theta - b_1 - b_2 f^1(\tilde{h}))^2$ and $\Pi^s_\ell = -\sum_{\theta\in\Theta}(\theta - f^1(\tilde{h}))^2$ are the agents silent payoffs. Thus, the statistic f^1 is essential for (u_e, u_ℓ). $\square$

Proof of Proposition 7.3. Fix an information structure. i.e. a set of public signals S and a mapping $\chi : \Theta \to S$. The prior μ and the information structure defines a probability space over $\Theta \times S$. Write $\mathbf{\Theta} : \Theta \times S \to \Theta$ for the realization of the state and $\mathbf{F} = \mathbb{E}[\mathbf{\Theta}|\mathcal{F}_\ell]$ for the layman's first-order expectation given the public signal. (Notice that $\mathcal{F}_\ell = \{\Theta \times \{s\} : s \in S\}$.) Write $c = \mathbb{E}[\mathbf{\Theta}]$. To show the result we first show some identities. Notice, by the law of iterated expectations,

$$\mathbb{E}[\mathbf{F}] = \mathbb{E}[\mathbb{E}[\mathbf{\Theta}|\mathcal{F}_\ell]] = \mathbb{E}[\mathbf{\Theta}] = c. \tag{16}$$

Since $\mathbf{F}$ is $\mathcal{F}_\ell$ measurable,

$$\mathbb{E}[\mathbf{F\Theta}] = \mathbb{E}[\mathbb{E}[\mathbf{F\Theta}|\mathcal{F}_\ell]] = \mathbb{E}[\mathbf{F}\mathbb{E}[\mathbf{\Theta}|\mathcal{F}_\ell]] = \mathbb{E}[\mathbf{F}^2], \tag{17}$$

where the first equality follows from the law of iterated expectations and the second by Theorem 4.1.14 in Durrett [2019]. Finally, note that

$$c^2 \leq \mathbb{E}[\mathbf{F}^2] \leq \mathbb{E}[\mathbf{\Theta}^2]. \tag{18}$$

where the fist inequality follows from Equation (16) and the second from the fact that

$$\mathbb{E}[(\mathbf{\Theta} - \mathbf{F})^2] = \mathbb{E}[\mathbf{\Theta}^2] - 2\mathbb{E}[\mathbf{\Theta F}] + \mathbb{E}[\mathbf{F}^2] = \mathbb{E}[\mathbf{\Theta}^2] - \mathbb{E}[\mathbf{F}^2].$$

Note, given the agent's payoffs π_e and π_ℓ, the ex ante total welfare is given by

$$\begin{aligned} TW &= \mathbb{E}[-(\mathbf{\Theta} - \mathbf{F})^2] + \mathbb{E}[-\lambda(\mathbf{F} - b_1 - b_2\mathbf{\Theta})^2] \\ &= \mathbb{E}[-\mathbf{\Theta}^2 + 2\mathbf{\Theta F} - \mathbf{F}^2] + \mathbb{E}[-\lambda \mathbf{F}^2 + 2\lambda b_1 + 2\lambda b_2 \mathbf{F\Theta} - \lambda(b_1 + b_2\mathbf{\Theta})^2] \\ &= \mathbb{E}[-\mathbf{\Theta}^2 + 2\mathbf{F}^2 - \mathbf{F}^2 - \lambda\mathbf{F}^2 + 2\lambda b_1 + 2\lambda b_2\mathbf{F}^2 - \lambda(b_1 + b_2\mathbf{\Theta})^2] \\ &= \mathbb{E}[-\mathbf{\Theta}^2 + 2\lambda b_1 - \lambda(b_1 + b_2\mathbf{\Theta})^2] + (1 + \lambda(2b_2 - 1)) \cdot \mathbb{E}[\mathbf{F}^2]. \end{aligned}$$

where the third equality follows from Equations (16) and (17). Notice that the first the term $\mathbb{E}[-\mathbf{\Theta}^2 + 2\lambda b_1 - \lambda(b_1 + b_2\mathbf{\Theta})^2]$ does not depend on the information structure used. Hence, the welfare maximizing information structure has to maximize $(1 + \lambda(2b_2 - 1))) \cdot \mathbb{E}[\mathbf{F}^2]$.

Notice, if $(1 + \lambda(2b_2 - 1)) \leq 0$, the information structure that maximizes welfare satisfies $\mathbb{E}[\mathbf{F}^2] = c^2$. (See Equation (18).) Hence, no information maximizes welfare. By contrast, if $(1 + \lambda(2b_2 - 1)) \geq 0$ the information structure that maximizes welfare satisfies $\mathbb{E}[\mathbf{F}^2] = \mathbb{E}[\mathbf{\Theta}^2]$. Hence, full revelation of the state maximizes total welfare. □

Proof of Proposition 7.4. The proof is divided in three steps. The first shows that (u_e, u_ℓ) is a reduced form for G. the second shows that $\text{RF} = \{(u_e, u_\ell)\}$ is a reduced-form representation of G. The third shows that $\overline{f}$ is essential for (u_e, u_ℓ).

Step 1. Fix a induced Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ and let $(\hat{\sigma}_e, \hat{\sigma}_\ell)$ be a Bayesian equilibrium thereof. First note that the strict concavity of π_i imply that each best response is single valued. (See Zimper [2006].) So, there is no loss to analyze only pure strategy profiles. Moreover, $(\hat{\sigma}_e, \hat{\sigma}_\ell)$

is a Bayesian equilibrium must satisfy the following first-order conditions:

$$\hat{\sigma}_e(\theta, T_e) = \theta + \alpha \sum_{T'_\ell \in \mathcal{T}_\ell} \hat{\sigma}_\ell(T'_\ell) \cdot \beta_e(\theta, T_e)(T'_\ell), \quad \text{and} \tag{19}$$

$$\hat{\sigma}_\ell(T_\ell) = \sum_{(\theta', T'_\ell) \in \Theta \times \mathcal{T}_\ell} \big(\theta' + \alpha \hat{\sigma}_e(\theta', T'_\ell)\big) \cdot \beta_\ell(T_e)(\theta', T'_e). \tag{20}$$

We show that the strategy profile $\hat{\sigma}_e$ and $\hat{\sigma}_\ell$ given by

$$\hat{\sigma}_e(\theta, T_e) = \theta + (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-1} \cdot \mathbb{E}_e^{2k} \theta(\delta_e(\theta, T_e)), \quad \text{and} \tag{21}$$

$$\hat{\sigma}_\ell(T_\ell) = (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} \cdot \mathbb{E}_\ell^{2k-1} \theta(\delta_\ell(T_\ell)) \tag{22}$$

satisfy Equations (19) and (20).

To show this, note that

$$\begin{aligned}
\hat{\sigma}_e(\theta, T_e) &= \theta + (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-1} \cdot \mathbb{E}_e^{2k} \theta(\delta_e(\theta, T_e)) \\
&= \theta + (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-1} \sum_{T'_\ell \in \mathcal{T}_\ell} \mathbb{E}_\ell^{2k-1} \theta(\delta_\ell(T'_\ell)) \cdot \beta_e(\theta, T_e)(T'_\ell) \\
&= \theta + \alpha \sum_{T'_\ell \in \mathcal{T}_\ell} (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} \cdot \mathbb{E}_\ell^{2k-1} \theta(\delta_\ell(T_\ell)) \cdot \beta_e(\theta, T_e)(T'_\ell) \\
&= \theta + \alpha \sum_{T'_\ell \in \mathcal{T}_\ell} \hat{\sigma}_\ell(T_\ell) \cdot \beta_e(\theta, T_e)(T'_\ell),
\end{aligned}$$

where the second equality follows from the fact that $\mathbb{E}_e^{2k}\theta(\cdot)$ is the expert's expectation of $\mathbb{E}_\ell^{2k-1}\theta(\cdot)$ and the last equality follows from the definition of $\hat{\sigma}_\ell$. So, Equation (19) holds.

In addition,

$$\begin{aligned}
\hat{\sigma}_\ell(T_\ell) &= (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-2} \mathbb{E}_\ell^{2k-1} \theta(\delta_\ell(T_\ell)) \\
&= (1+\alpha) \sum_{(\theta', T'_e) \in \Theta \times \mathcal{T}_e} \left(\theta + \sum_{k=1}^{\infty} \alpha^{2k} \mathbb{E}_e^{2k} \theta(\delta_e(\theta', T'_\ell)\right) \cdot \beta_\ell(T_\ell)(\theta', T'_e) \\
&= \sum_{(\theta', T'_\ell) \in \Theta \times \mathcal{T}_\ell} \left(\theta' + \alpha\theta' + (1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k} \mathbb{E}_e^{2k} \theta(\delta_e(\theta', T'_\ell)\right) \cdot \beta_\ell(T_\ell)(\theta', T'_e) \\
&= \sum_{(\theta', T'_\ell) \in \Theta \times \mathcal{T}_\ell} \left(\theta' + \alpha\theta' + \alpha(1+\alpha) \sum_{k=1}^{\infty} \alpha^{2k-1} \mathbb{E}_e^{2k} \theta(\delta_e(\theta', T'_\ell)\right) \cdot \beta_\ell(T_\ell)(\theta', T'_e) \\
&= \sum_{(\theta', T'_\ell) \in \Theta \times \mathcal{T}_\ell} (\theta' + \alpha \hat{\sigma}_e(\theta', T'_\ell)) \cdot \beta_\ell(T_e)(\theta', T'_e).
\end{aligned}$$

where the second equality follows from the fact that $\mathbb{E}_\ell^1\theta(\cdot)$ and $\mathbb{E}_\ell^{2k+1}\theta(\cdot)$ are the layman's expectation of θ and $\mathbb{E}_\ell^{2k}\theta(\cdot)$ respectively. Hence, Equation (20) also holds and $(\hat{\sigma}_e, \hat{\sigma}_\ell)$ is a Bayesian equilibrium.

Finally, notice that

$$\begin{aligned}\Pi_e(\theta, T_e \mid \hat{\sigma}, \beta) &= \hat{\sigma}_e(\theta, T_e)\left(\theta - \tfrac{1}{2}\hat{\sigma}_e^2(\theta, T_e) + \alpha \textstyle\sum_{T_\ell \in \mathcal{T}_e} \hat{\sigma}_\ell(T_\ell) \cdot \beta_e(\theta, T_e)(T_\ell)\right) \\ &= \hat{\sigma}_e(\theta, T_e)\left(\tfrac{1}{2}\hat{\sigma}_e^2(\theta, T_e)\right) \\ &= \tfrac{1}{2}\hat{\sigma}_e(\theta, T_e)^2 \\ &= u_e(\theta, \delta_e(\theta, T_e)).\end{aligned}$$

where the second equality follows from Equation (19).

$$\begin{aligned}\Pi_\ell(T_\ell \mid \hat{\sigma}, \beta) &= \hat{\sigma}_\ell(T_\ell)\left(\textstyle\sum_{(\theta, T_e) \in \Theta \times \mathcal{T}_e} \theta - \tfrac{1}{2}\hat{\sigma}_\ell(T_\ell) + \alpha \hat{\sigma}_e(T_e) \cdot \beta_\ell(T_\ell)(\theta, T_e)\right) \\ &= \hat{\sigma}_e(T_\ell)\left(\tfrac{1}{2}\hat{\sigma}_e^2(T_\ell)\right) \\ &= \tfrac{1}{2}\hat{\sigma}_e(T_\ell)^2 \\ &= u_\ell(\delta_\ell(T_\ell)).\end{aligned}$$

where the second equality follows from Equation (20). Thus, (u_e, u_ℓ) is a reduced for for G.

Step 2. Fix an induced Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. By Lemma B.13, it suffices to show that $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ has a unique BE. We show that the strategy profile $\hat{\sigma} = (\hat{\sigma}_e, \hat{\sigma}_\ell)$ given by Equations (21) and (22) is the unique Bayesian equilibrium of $(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$.

Fix a Bayesian equilibrium $\tilde{\sigma}$ of the induced Bayesian game. We will show that $\tilde{\sigma} = \hat{\sigma}$. Since $\tilde{\sigma} = (\tilde{\sigma}_e, \tilde{\sigma}_\ell)$ is a Bayesian equilibrium, then it satisfies the first order conditions

$$\tilde{\sigma}_e(\theta, T_e) = \theta + \alpha \sum_{T'_\ell \in \mathcal{T}_\ell} \sigma_\ell(T'_\ell) \cdot \beta_e(\theta, T_e)(T'_\ell) \quad \text{and} \tag{23}$$

$$\tilde{\sigma}_\ell(T_\ell) = \sum_{(\theta', T'_\ell) \in \Theta \times \mathcal{T}_\ell} (\theta + \alpha \sigma_e(\theta, T_\ell)) \cdot \beta_\ell(T_\ell)(\theta', T'_e). \tag{24}$$

Thus, by Equations (19) and (23),

$$\begin{aligned}\sup_{\theta, T_e} |\hat{\sigma}_e(\theta, T_e) - \tilde{\sigma}_e(\theta, T_e)| &= \left|\alpha \textstyle\sum_{T_\ell \in \mathcal{T}_\ell} (\hat{\sigma}_\ell(T_\ell) - \tilde{\sigma}_\ell(T_\ell)) \cdot \beta_e(\theta, T_e)(T_\ell)\right| \\ &\leq |\alpha| \cdot \sup_{T_\ell \in \mathcal{T}_\ell} |\hat{\sigma}_\ell(T_\ell) - \tilde{\sigma}_\ell(T_\ell)|.\end{aligned}$$

Similarly, by Equations (20) and (23),

$$\begin{aligned}\sup_{T_\ell} |\hat{\sigma}_\ell(T_\ell) - \tilde{\sigma}_e(T_\ell)| &= \left|\alpha \textstyle\sum_{(\theta, T_e) \in \Theta \times \mathcal{T}_e} (\hat{\sigma}_\ell(T_\ell) - \tilde{\sigma}_\ell(T_\ell)) \cdot \beta_\ell(T_\ell)(\theta, T_e)\right| \\ &\leq |\alpha| \cdot \sup_{(\theta, T_e) \in \Theta \times \mathcal{T}_e} |\hat{\sigma}_e(\theta, T_\ell) - \tilde{\sigma}_e(\theta, T_\ell)|.\end{aligned}$$

Since $|\alpha| < 1$, the two inequalities above imply that

$$\sup_{(\theta,T_e)\in\Theta\times\mathcal{T}_e} |\hat{\sigma}_e(\theta,T_e) - \tilde{\sigma}_e(\theta,T_e)| = \sup_{T_\ell\in\mathcal{T}_\ell} |\hat{\sigma}_\ell(T_\ell) - \tilde{\sigma}_e(T_\ell)| = 0.$$

So $(\tilde{\sigma}_e, \tilde{\sigma}_\ell) = (\hat{\sigma}_e, \hat{\sigma}_\ell)$, an the induced Bayesian game $(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ has a unique BE.

Step 3. Fix a credible direct mechanism $\mathcal{M}^d$ and let ϕ the associated ex-ante probability measure it induces. Write (ρ^*, β^*) for the honest profile. If $\mathcal{M}^d$ is not informative about $\overline{f}$, it follows that $\overline{\mathbf{F}} = \overline{f}(\tilde{h})$ ϕ-almost surely. Hence,

$$\begin{aligned}\mathcal{U}_e(\rho^*|, \theta, \{\emptyset\}, \beta^*) =& \Pi^s_e(\theta) + \mathbb{E}Y_e(\theta|\mathcal{M}^d, \sigma^*), \text{ and}\\ \mathcal{U}_\ell(\rho^*|\{\emptyset\}, \beta^*) =& \Pi^s_\ell + \mathbb{E}Y_\ell(\theta|\mathcal{M}^d, \sigma^*).\end{aligned}$$

where, $\Pi^s_e(\theta) = \frac{1}{2}(\theta + \alpha\overline{f}(\hat{h}))^2$ and $\Pi^s_\ell = \frac{1}{2}\sum_{\theta\in\Theta}(\theta + \alpha\overline{f}(\hat{h}))^2$ are the agent's silent payoffs. Thus, the statistic $\overline{f}$ is essential for (u_e, u_ℓ). □

Proposition B.2. *Consider the game G of Section 7.2. In comparison with no information sharing, the layman is strictly better off with full revelation of the state. Moreover, fully revealing the state increases utilitarian welfare if and only if $\alpha \in (1-\sqrt{2}, 1)$.*

Proof. We use the reduced form (u_e, u_ℓ) found in Proposition 7.5 to compute the agents' ex ante utility under full revelation of the state and no information sharing.

Write $b = \mathbb{E}[\mathbf{\Theta}]$ and $d = (1+\alpha)\sum_{k=1}^{\infty} \alpha^{2k} = (1-\alpha)^{-1}$. Note that $1 + \alpha d = d$. Notice also that $\mathbb{E}_i\overline{f}(h_i^\theta) = d\theta$ and $\mathbb{E}_i\overline{f}(\widetilde{h}_i) = db$. Write FR_i (resp. NI_i) for the expected utility under full revelation (resp. no information) for agent i.

The ex ante expert's benefit from information sharing is

$$\begin{aligned}\mathrm{FR}_e - \mathrm{NI}_e &= \mathbb{E}[(\mathbf{\Theta} + \alpha d\mathbf{\Theta})^2] - \mathbb{E}[(\mathbf{\Theta} + \alpha db)^2]\\ &= \mathrm{Var}[\mathbf{\Theta} + \alpha d\mathbf{\Theta}] - \mathrm{Var}[\mathbf{\Theta} + \alpha db]\\ &= ((1+\alpha d)^2 - 1)\mathrm{Var}[\mathbf{\Theta}]\\ &= (d^2 - 1)\mathrm{Var}[\mathbf{\Theta}],\end{aligned}$$

where the second equality follows from the fact that $(\mathbb{E}[\mathbf{\Theta} + \alpha db])^2 = (\mathbb{E}[\mathbf{\Theta} + \alpha d\mathbf{\Theta}])^2$, and the last equality follows from the fact that $1 + \alpha d = d$. Notice that the expert gets better off with information sharing only if $d = (1-\alpha)^{-1} \geq 1$, i.e., only if $\alpha \geq 0$.

The ex ante layman's benefit from information sharing is

$$\begin{aligned}\mathrm{FR}_\ell - \mathrm{NI}_\ell &= \mathbb{E}[(d\mathbf{\Theta})^2] - \mathbb{E}[(db)^2]\\ &= \mathrm{Var}[(d\mathbf{\Theta})^2] - \mathrm{Var}[(db)^2]\\ &= d^2\mathrm{Var}[\mathbf{\Theta}^2],\end{aligned}$$

where the second equality follows form the fact that $(\mathbb{E}[d\mathbf{\Theta}])^2 = (\mathbb{E}[db])^2$.

Notice that this value is always positive. Hence, irrespectively of α, the layman gets positive benefits from information sharing. Notice that information sharing increases the agents' total welfare if and only if $2d^2 - 1 \geq 0$. Since $d^2 = (1-\alpha)^2 > 0$ this is equivalent to $\alpha \in (1-\sqrt{2}, 1)$ as desired. $\square$

B.5 Proofs from Section 8

Lemma B.13. *Consider a set of finite type structures $\mathcal{S}$, $(S_e, S_\ell, b_e, b_\ell)$, where (i) each $S_i \subseteq \mathbb{R}$ is finite and (ii) $b_e : \Theta \times S_e \to \Delta(S_\ell)$ and $b_\ell : S_\ell \to \Delta(\Theta \times S_e)$ are belief maps. If, for each type structure in $\mathcal{S}$, the associated Bayesian game $(G, \mathcal{S})$ has a unique Bayesian equilibrium, then*

(i) G has a reduced form (u_e, u_ℓ), and

(ii) for each reduced form (u_e, u_ℓ), $\mathrm{RF} = \{(u_e, u_\ell)\}$ is a reduced-form representation for G.

The proof of Lemma B.13 require some definitions. Say $h_i \in H_i$ is a **feasible** if there is a finite hierarchy set $M_e \times M_\ell \subsetneq H_e \times H_\ell$ such that $h_i \in M_i$ and $M_e \times M_\ell$ is belief closed. Write $\overline{H}_i$ for the set of feasible hierarchies of beliefs. Notice if $M_e \times M_\ell$ and $M_e' \times M_\ell'$ are both belief closed sets and their interception is non empty, then $(M_e \cap M_\ell)' \times (M_e \cap M_\ell)'$ is also belief closed. Hence, for each $h_i \in \overline{H}_i$, there is a finite belief-closed set $M_e \times M_\ell$ with $h_i \in M_i$, such that for each other belief closed set $M_e' \times M_\ell'$ such that $h_i \in M_i'$, it follows that $M_e \times M_\ell \subseteq M_e' \times M_\ell'$. Call such set $M_e \times M_\ell$ the hierarchy set **generated** by h_i.

Proof of Lemma B.13.
Part (i). The proof first defines belief-based utility functions (u_e, u_ℓ) and then it shows that (u_e, u_ℓ) is a reduced form for G.

Start with the definition of u_e. Fix $(\theta, h_e) \in \Theta \times H_e$. If $h_e \notin \overline{H}_e$, set $u_e(\theta, h_e) = 0$. If $h_e \in \overline{H}_e$, we will define $u_e(\theta, h_e)$ as the expected payoff of a Bayesian equilibrium of an induced Bayesian game we will construct. To do so, let $M_e \times M_\ell \subsetneq H_e \times H_\ell$ be the hierarchy set generated by h_e and let $\mathcal{M}^d = ((Y_i, M_i) : i \in \{e, \ell\}, m)$ with $Y_e = Y_\ell = \{0\}$. (The choice of m is irrelevant). Recall that $T_e^d[\theta, h_e] = \{\theta\} \times \{h_e\} \times M_\ell \times Y$ is the expert's terminal information set associated with report θ and hierarchy-message h_e. Likewise, recall that $T_\ell^d[h_\ell] = \Theta \times M_e \times \{h_\ell\} \times Y$ is the layman's terminal information set associated with hierarchy-message h_ℓ. Attach to the induced bayesian game straight forward beliefs β^*. So, the associated hierarchy mappings satisfy $\delta_e^*(\theta, T_e^d[\theta, h_e]) = h_e$ and $\delta_\ell^*(T_\ell[h_\ell]) = h_\ell$ for each $(\theta, h_e, h_\ell) \in \Theta \times M_e \times H_\ell$. (See Lemma B.1.) Notice $BG(\mathcal{M}^d, \beta^*)$ has finite sets of types. Hence, by assumption, the induced Bayesian game $BG(\mathcal{M}^d, \beta^*)$ has a unique Bayesian equilibrium $\hat{\sigma}$. Set $u_e(\theta, h_e) = \Pi_e(\hat{\sigma} | \theta, T_e^d[\theta, h_e])$.

We now define u_ℓ. Fix $h_\ell \in H_\ell$. If $h_\ell \notin \overline{H}_\ell$ set $u_\ell(h_\ell) = 0$. If $h_\ell \in \overline{H}_\ell$, let $M_e \times M_\ell \subsetneq H_e \times H_\ell$ the hierarchy generated by h_ℓ and construct the induced Bayesian game $BG(\mathcal{M}^d, \beta^*)$ in an analogous way as above. Let $\hat{\sigma}$ the unique Bayesian equilibrium of $BG(\mathcal{M}^d, \beta^*)$ and set $u_\ell(h_\ell) = \Pi_\ell(\hat{\sigma} | \theta, T_\ell^d[h_\ell])$.

We now show that (u_e, u_ℓ) is a reduced form. Notice, by Proposition B.1 it suffices to show that for each credible direct mechanism $\mathcal{M}^d = ((Y_i, M_i) : i \in \{e, \ell\}, m)$ with $Y_e = Y_\ell = \{0\}$ and each straight forward beliefs β^*, there is a Bayesian equilibrium $\hat{\sigma}$ of $BG(\mathcal{M}^d, \beta^*)$ such that

(1) For each $(\theta, \theta', h_e) \in \Theta \times \Theta \times M_e$, $u_e(\theta, h_e) = \Pi_e(\hat{\sigma}|\theta, T_e^d[\theta', h_e], \beta_e^*)$.

(2) For each $h_\ell \in M_\ell$, $u_\ell(h_\ell) = \Pi_\ell(\hat{\sigma}|T_\ell^d[\theta, h_\ell], \beta_\ell^*)$.

We show (1) by contradiction. (Showing (2) follows from an analogous argument). Assume (1) is not satisfied. That is, that there are some $(\theta, \theta', h_e) \in \Theta \times \Theta \times M_e$ such that $u_e(\theta, h_e) \neq \Pi_e(\hat{\sigma}|\theta, T_e^d[\theta', h_e], \beta_e^*)$. Let $\underline{M}_e \times \underline{M}'_\ell$ be the hierarchy set induced by h_e and notice that $\underline{M}_e \times \underline{M}_\ell \subseteq M_e \times M_\ell$. Hence, The set $\underline{M}_e \times \underline{M}_\ell$ induces a "smaller" direct mechanism $\underline{\mathcal{M}}^d = ((Y_i, M_i) : i \in \{e, \ell\}, m)$ with $Y_e = Y_\ell = \{0\}$. (the protocol m is irrelevant). Write $\underline{\beta}^*$ for the straight-forward beliefs associated to $\underline{\mathcal{M}}^d$ and note that for each $(\theta, \theta', h_e, h_\ell) \in \Theta \times \Theta \times \underline{M}_e \times \underline{M}_\ell$,

$$\beta_e^*(\theta, T_e^d[\theta', h_e])(T_\ell^d[h_\ell]) = \underline{\beta}_e^*(\theta, T_e^d[\theta', h_e])(T_\ell^d[h_\ell]), \quad \text{and} \tag{25}$$

$$\beta_\ell^*(T_\ell^d[h_\ell])(T_e^d[\theta, h_e]) = \underline{\beta}_\ell^*(T_\ell^d[h_\ell])(T_e^d[\theta, h_e]). \tag{26}$$

Notice that the equilibrium $\hat{\sigma} = (\hat{\sigma}_e, \hat{\sigma}_\ell)$ induces a strategy profile $(\underline{\hat{\sigma}}_e, \underline{\hat{\sigma}}_\ell)$ of the "smaller" Bayesian equilibrium $BG(\underline{\mathcal{M}}^d, \underline{\beta}^*)$ defined by restricting $\hat{\sigma}_e$ and $\hat{\sigma}_\ell$ to the information sets of $BG(\underline{\mathcal{M}}^d, \underline{\beta}^*)$. More specifically, for each $(\theta, \theta', h_e, h_\ell) \in \Theta \times \Theta \times \underline{M}_e \times \underline{M}_\ell$,

$$\underline{\sigma}_e^*(\theta, T_e^d[\theta', h_e]) = \sigma_e^*(\theta, T_e^d[\theta', h_e]), \quad \text{and}$$
$$\underline{\sigma}_\ell^*(T_\ell^d[h_\ell]) = \sigma_\ell^*(T_\ell^d[h_\ell]).$$

Note, since $\underline{\beta}_e$ satisfies Equations (25) and (26) it follows that the agents get the same expected payoffs under $\underline{\hat{\sigma}}$ in $BG(\underline{\mathcal{M}}^d, \underline{\beta})$ than under $\hat{\sigma}$ in $BG(\mathcal{M}^d, \beta)$. More specifically,

$$\Pi_e(\underline{\hat{\sigma}}|\theta, T_e^d[\theta', h_e], \underline{\beta}_e^*) = \Pi_e(\hat{\sigma}|\theta, T_e^d[\theta', h_e], \beta_e^*), \quad \text{and}$$
$$\Pi_\ell(\underline{\hat{\sigma}}|\theta, T_e^d[\theta', h_e], \underline{\beta}_\ell^*) = \Pi_e(\hat{\sigma}|\theta, T_e^d[\theta', h_e], \beta_\ell^*).$$

Moreover, since $\hat{\sigma}$ is an equilibrium in $BG(\mathcal{M}^d, \beta^*)$ it follows that $\underline{\hat{\sigma}}$ is an equilibrium in $BG(\underline{\mathcal{M}}^d, \underline{\beta}^*)$. However, by uniqueness of equilibria of $BG(\mathcal{M}^d, \beta^*)$ and by construction of u_e, for each $(\theta, \theta', h_e) \in \Theta \times \Theta \times M_e$,

$$u_e(\theta, h_e) = \Pi_e(\underline{\hat{\sigma}}|\theta, T_e^d[\theta', h_e], \beta_e^*) = \Pi_e(\hat{\sigma}|\theta, T_e^d[\theta', h_e], \beta_e^*),$$

which leads to a contradiction. Hence (u_e, u_ℓ) is a reduced form of G.

Part (ii). Fix a mechanism $\mathcal{M} = (\cdot, \mathcal{T}_e, \mathcal{T}_\ell)$ and a PBE (σ, β) of a supergame $(\mathcal{M}, G)$. Let ρ be the induced strategy in the psychological game $(\mathcal{M}, u_e, u_\ell)$. We show that (ρ, β) is a PBE of $(\mathcal{M}, u_e, u_\ell)$ that is payoff equivalent to (σ, β).

Let $\hat{\sigma}$ be the unique Bayesian equilibrium of the induced Bayesian game $(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. Since (σ, β) is a PBE of $(\mathcal{M}, G)$ and BG$(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ has a unique Bayesian equilibrium, then σ must induce $\hat{\sigma}$. Otherwise, (σ, β) does not satisfies sequential rationality at the terminal information sets $\mathcal{T}_e$ and $\mathcal{T}_\ell$. Write ρ for the strategy that σ induces in the psychological game $(\mathcal{M}, u_e, u_\ell)$.

First we show payoff equivalence for the expert. Notice that for each $\theta \in \Theta$,

$$\begin{aligned}\mathcal{U}_e(\rho \mid \theta, \emptyset, \beta) &= \sum_{T_e \in \mathcal{T}_e} [u_e(\theta, \delta_e(\theta, T_e)) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid \rho, \theta, \emptyset, \beta_e) \\ &= \sum_{T_e \in \mathcal{T}_e} [\Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid \rho, \theta, \emptyset, \beta_e) \\ &= U_e(\sigma \mid \theta, \emptyset, \beta),\end{aligned}$$

where the second equality follows from the fact that $\hat{\sigma}$ is the BE associated to (u_e, u_ℓ). Moreover, $\mathbb{E}Y_e(\theta \mid \mathcal{M}, \sigma, \beta_e) = \mathbb{E}Y_e(\theta \mid \mathcal{M}, \rho, \beta_e)$.

Using an analogous argument, for each $I_\ell \in \mathcal{I}_\ell$, $U_\ell(\sigma \mid I_\ell, \beta) = \mathcal{U}_\ell(\rho \mid I_\ell, \beta)$ and $\mathbb{E}Y_\ell(\mathcal{M}, \sigma, \beta_e) = \mathbb{E}Y_\ell(\mathcal{M}, \rho, \beta_e)$. Hence, (σ, β) and (ρ, β) are payoff equivalent.

It suffices to show that (ρ, β) is a perfect Bayesian equilibrium of $(\mathcal{M}, u_e, u_\ell)$. Notice that β is consistent with ρ, so β is also consistent with σ. Fix a strategy ρ'_e for the expert in the supergame and let σ'_e the supergame's strategy induced by $(\rho'_e, \hat{\sigma}_e)$. Then,

$$\begin{aligned}\mathcal{U}_e(\rho \mid \theta, I_e, \beta) &= U_e(\sigma \mid \theta, I_e, \beta) \\ &\geq U_e(\sigma'_e, \sigma_\ell \mid \theta, I_e, \beta), \\ &= \sum_{T_e \in \mathcal{T}_e} [\Pi_e(\hat{\sigma} \mid \theta, T_e, \beta_e) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid (\rho'_e, \rho_\ell), \theta, I_e, \beta_e) \\ &= \sum_{T_e \in \mathcal{T}_e} [u_e(\theta, \delta_e(\theta, T_e)) + \gamma_e(T_e)] \cdot \mathbb{P}(T_e \mid (\rho'_e, \rho_\ell), \theta, I_e, \beta_e) \\ &= \mathcal{U}_e(\rho'_e, \rho_\ell \mid \theta, I_e, \beta),\end{aligned}$$

where the inequality follows from the fact that (σ, β) is a PBE of $(\mathcal{M}, G)$. An analogous argument establishes that for each strategy ρ'_ℓ and each information set $I_\ell \in \mathcal{I}_\ell$, it follows that $U_\ell(\sigma \mid I_\ell, \beta) \geq U_\ell(\sigma_e, \sigma'_\ell \mid I_\ell, \beta)$. So, sequential rationality is satisfied and (ρ, β) is a PBE of the $(\mathcal{M}, u_e, u_\ell)$. □

The second existence result requires the introduction of some notation. Say that the **expert is inactive** in $G = ((A_i, \pi_i) : i \in \{e, \ell\})$ if A_e is a singleton. For notational simplicity, write $\pi_i : \Theta \times A_\ell \to \mathbb{R}$ for the agents payoffs and write $\pi_\ell(\mu^1_\ell, a_\ell) = \sum_{\theta \in \Theta} \pi_\ell(\theta, a_\ell) \cdot \mu^1_\ell(\theta)$ for the layman's expected payoff of a given action and a given first-order belief. In addition, write

$$Q(G) = \left\{ q : \Delta(\Theta) \to \Delta(A_\ell) : \text{Supp}\,(q(\mu^1_\ell)) \subseteq \arg\max_{A_\ell} \pi_\ell(\mu^1_\ell, a_\ell) \text{ for each } \mu^1_\ell \in \Delta(\Theta) \right\},$$

for the optimal layman's plans contingent on his first-order beliefs. Finally, say that the belief-based utilities (u_e, u_ℓ) are induced by $q \in Q(G)$ if

$$\begin{aligned}u_\ell(\mu^1_\ell, \mu^2_\ell, ...) &= \int_{A_\ell} \pi_\ell(\mu^1_\ell, a_\ell)\ dq(\mu^1_\ell), \text{ and} \\ u_e(\theta, \mu^1_e, \mu^2_e, ...) &= \int_{\Delta(\Theta)} \int_{A_\ell} \pi_e(\theta, a_\ell)\ dq(\mu^1_\ell)\ d\mu^2_e.\end{aligned}$$

The following lemma provides an existence result for games where the expert is inactive.

Lemma B.14. *Suppose that the expert is inactive in G and $Q(G) \neq \emptyset$. Then, the following holds:*

(i) If (u_e, u_ℓ) is induced by $q \in Q(G)$, then (u_e, u_ℓ) is a reduced form for G.

(ii) The set RF $= \{(u_e, u_\ell) : (u_e, u_\ell)$ *is induced by some* $q \in Q(G)\}$ *is a reduced form representation for* G.

The proof of B.14 requires the introduction of some notation and Lemma B.15 below. Fix a credible direct mechanism $\mathcal{M}^d = ((Y_i, M_i, i \in \{e, \ell\}), m)$. Write $M_i^k = \text{proj}_{\Delta(D_i^k)}(M_i)$ for the set of k-order beliefs of the set M_i, and for each $\mu_i^k \in M_i^k$, write $M_i^k[\mu_i^k] = \{h_i \in M_i : \text{proj}_{\Delta(D_i^k)}(h_i) = \mu_i^k\}$ for the set of hierarchy messages in M_i^k with k-order belief μ_i^k.

Lemma B.15. *Let* G *be a game with an inactive expert. Fix a credible direct mechanism* $\mathcal{M}^d = ((M_i, Y_i : i \in \{e, \ell\}))$ *and* (u_e, u_ℓ) *belief-based utility functions induced by* $q \in Q(G)$. *Then, under* (u_e, u_ℓ) *the agents value of* $\mathcal{M}^d$ *is given by:*

(i) $V_e(\theta, \theta'|\mathcal{M}^d) = \sum_{\mu_\ell^1 \in M_\ell^1} \left(\int_{A_\ell} \pi_\ell(\theta, a_e)\ q(\mu_\ell^1) \right) \cdot \text{marg}_{M_\ell} m(\theta')(M_\ell^1[\mu_\ell^1]) + \mathbb{E}Y_e(\theta'|\mathcal{M}^d, \rho^*).$

(ii) $V_\ell(\mathcal{M}^d) = \sum_{\mu_\ell^1 \in M_\ell^1} \sum_{\theta \in \Theta} \left(\int_{A_\ell} \pi_\ell(\theta, a_e)\ q(\mu_\ell^1) \right) \cdot \text{marg}_{\Theta \times M_\ell} \phi(\{\theta\} \times M_\ell^1[\mu_\ell^1]) + \mathbb{E}Y_\ell(\mathcal{M}^d, \rho^*).$

Proof. Fix $q \in Q(G)$. We will compute the agents' value of $\mathcal{M}^d$ for psychological payoffs (u_e, u_ℓ). First, we compute the expert's value of $\mathcal{M}^d$. By credibility, for each $(\theta, h_\ell) \in \Theta \times M_\ell[\mu_\ell^1]$,

$$\sum_{h_e \in M_e} \sum_{y_\ell \in Y_\ell} \eta_\ell(h_\ell)(\theta, h_e) \cdot \text{marg}_{Y_\ell \times M_\ell} \phi(y_\ell, h_\ell) = \sum_{h_e \in M_e} \sum_{y_\ell \in Y_\ell} \text{marg}_{\Theta \times M_e \times Y_\ell \times M_\ell} \phi(\theta, h_e, y_\ell, h_\ell),$$

which implies

$$\mu_\ell^1(\theta) \cdot \text{marg}_{M_\ell} \phi(h_\ell) = \text{marg}_{\Theta \times M_\ell} \phi(\theta, h_\ell). \tag{27}$$

So, the value of $\mathcal{M}^d$ for the layman is given by

$$\begin{aligned}
V_\ell(\mathcal{M}^d) &= \sum_{\mu_\ell^1 \in M_\ell^1} \sum_{h_\ell \in M_\ell[\mu_\ell^1]} u_\ell(h_\ell) \cdot \text{marg}_{M_\ell} \phi(h_\ell) + \mathbb{E}Y_\ell(\mathcal{M}^d, \rho^*) \\
&= \sum_{\mu_\ell^1 \in M_\ell^1} \sum_{h_\ell \in M_\ell[\mu_\ell^1]} \sum_{\theta \in \Theta} \left(\int_{A_\ell} \pi_\ell(\theta, a_e)\ q(\mu_\ell^1) \right) \cdot \mu_\ell^1(\theta) \cdot \text{marg}_{M_\ell} \phi(h_\ell) + \mathbb{E}Y_\ell(\mathcal{M}^d, \rho^*) \\
&= \sum_{\mu_\ell^1 \in M_\ell^1} \sum_{\theta \in \Theta} \left(\int_{A_\ell} \pi_\ell(\theta, a_e)\ q(\mu_\ell^1) \right) \sum_{h_\ell \in M_\ell[\mu_\ell^1]} \text{marg}_{\Theta \times M_\ell} \phi(\theta, h_\ell) + \mathbb{E}Y_\ell(\mathcal{M}^d, \rho^*) \\
&= \sum_{\mu_\ell^1 \in M_\ell^1} \sum_{\theta \in \Theta} \left(\int_{A_\ell} \pi_\ell(\theta, a_e)\ q(\mu_\ell^1) \right) \cdot \text{marg}_{\Theta \times M_\ell} \phi(\{\theta\} \times M_\ell^1[\mu_\ell^1]) + \mathbb{E}Y_\ell(\mathcal{M}^d, \rho^*),
\end{aligned}$$

where the second equality follows from definition of u_e, and the third from Equation (27).

Now we compute the expert's value of $\mathcal{M}^d$. By credibility, for each $(\theta, h_e) \in \Theta \times M_e[\mu_e^2]$,

$$\sum_{h_\ell \in M_\ell[\mu_\ell^1]} \sum_{y_e \in Y_e} \eta_e(h_e)(h_\ell) \cdot \text{marg}_{\Theta \times M_e \times Y_e} \phi(\theta, h_e, y_e) = \sum_{h_\ell \in M_\ell[\mu_\ell^1]} \sum_{y_e \in Y_e} \text{marg}_{\Theta \times M_e \times Y_e \times M_\ell} \phi(\theta, h_e, y_e, h_\ell),$$

which implies

$$\mu_e^2(\mu_\ell^1) \cdot \text{marg}_{\Theta \times M_\ell} \phi(\theta, h_e) = \text{marg}_{\Theta \times M_e \times M_\ell} \phi(\{(\theta, h_e)\} \times M_\ell^1[\mu_\ell^1]).$$

Then, since $m(\theta)(y,h)\cdot\mu(\theta)=\phi(\theta,y,h)$ and $\mu(\theta)\neq 0$,

$$\mu_e^2(\mu_\ell^1)\cdot \text{marg}_{M_\ell} m(\theta)(h_e)=\text{marg}_{M_e\times M_\ell} m(\theta)(\{h_e\}\times M_\ell^1[\mu_\ell^1]). \tag{28}$$

Hence, the expert's value of reporting θ' when state is θ in $\mathcal{M}^d$ is

$$\begin{aligned}
V_e(\theta,\theta'|\mathcal{M}^d) &= \sum_{\mu_e^2\in M_e^2}\sum_{h_e\in M_e[\mu_e^2]} u_e(\theta,h_e)\cdot \text{marg}_{H_e} m(\theta')(h_e)+\mathbb{E}Y_e(\theta'|\mathcal{M}^d,\rho^*)\\
&= \sum_{\mu_e^2\in M_e^2}\sum_{h_e\in M_e[\mu_e^2]}\sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_e(\theta,a_\ell)\;dq(\mu_\ell^1)\right)\cdot\mu_e^2(\mu_\ell^1)\cdot\text{marg}_{H_e} m(\theta')(h_e)+\mathbb{E}Y_e(\theta'|\mathcal{M}^d,\rho^*)\\
&= \sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_e(\theta,a_\ell)\;dq(\mu_\ell^1)\right)\sum_{\mu_e^2\in M_e^2}\sum_{h_e\in M_e[\mu_e^2]}\cdot\mu_e^2(\mu_\ell^1)\cdot\text{marg}_{H_e} m(\theta')(h_e)+\mathbb{E}Y_e(\theta'|\mathcal{M}^d,\rho^*)\\
&= \sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_e(\theta,a_\ell)\;dq(\mu_\ell^1)\right)\sum_{\mu_e^2\in M_e^2}\sum_{h_e\in M_e[\mu_e^2]}\text{marg}_{H_e\times H_\ell} m(\theta')(\{h_e\}\times M_\ell^1[\mu_\ell^1])+\mathbb{E}Y_e(\theta'|\mathcal{M}^d,\rho^*)\\
&= \sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_\ell(\theta,a_e)\;q(\mu_\ell^1)\right)\cdot\text{marg}_{M_\ell} m(\theta')(M_\ell^1[\mu_\ell^1])+\mathbb{E}Y_e(\theta'|\mathcal{M}^d,\rho^*),
\end{aligned}$$

where the second equality follows from definition of u_ℓ, and the third from Equation (28). □

Proof of Lemma B.14. First We show (i). Fix a mechanism $\mathcal{M}=(\cdot,\mathcal{T}_e,\mathcal{T}_\ell)$ and some $\beta\in\text{Conv}(\mathcal{M})$ that induces (δ_e,δ_ℓ). Consider the induced Bayesian game $BG(\mathcal{T}_e,\mathcal{T}_\ell,\beta)$. Define the strategy profile $\hat\sigma$ of the induced Bayesian game by $\hat\sigma(T_\ell)=q(\delta_\ell^1(T_\ell))$. (Note, the expert is inactive.) Since q is optimal, it follows that $\hat\sigma$ is a Bayesian equilibrium. Moreover, by construction, for each $T_\ell\in\mathcal{T}_\ell$, $\Pi_\ell(\hat\sigma|T_\ell,\beta_\ell)=u_\ell(\delta_\ell(T_\ell))$. Similarly, for each $(\theta,T_e)\in\Theta\times\mathcal{T}_e$, $\Pi_e(\hat\sigma|\theta,T_e,\beta_e)=u_e(\theta,\delta_e(\theta,T_e))$.

Now we show (ii). Fix an individually rational PBE (σ,β) of $(\mathcal{M},G)$ and let $\mathcal{M}^d$ be the extended direct mechanism induced by (σ,β). We will show that there is an optimal contingent plan $q\in Q$ such that (σ,β) is equivalent to each honest strategy profile (ρ^*,β^*) of the psychological game $(\mathcal{M}^d,u_e,u_\ell)$.

First we select q. For each posterior $\mu_\ell^1\in M_\ell^1$ write $Z[\mu_\ell^1]=\{z\in Z:\delta_\ell^1(T_\ell[z])=\mu_\ell^1]\}$. Notice that, since (σ,β) is a PBE, $\delta_\ell^1(T_\ell)=\mu_\ell^1$ implies $\text{Supp}\,(\sigma_\ell(T_\ell))\subseteq\arg\max_{A_\ell}\pi_\ell(\mu_\ell^1,a_\ell)$. Thus, there is some optimal contingent plan $q\in Q(G)$ such that for each $\mu_\ell^1\in M_\ell^1$,

$$q(\mu_\ell^1)\cdot\mathbb{P}(Z[\mu_\ell^1]|\sigma)=\sum_{z\in Z[\mu_\ell^1]}\sigma_\ell(T_\ell[z])\cdot\mathbb{P}(z|\sigma). \tag{29}$$

So, for each posterior $\mu_\ell^1\in M_\ell^1$, $q(\mu_\ell^1)$ is the expected layman's strategy conditional on the set $Z[\mu_\ell^1]$.

Fix $\theta,\theta'\in\Theta$. To show equivalence of (σ,β) and (ρ^*,β^*) it suffices to show that there is some expert's strategy σ_e' such that

$$\mathcal{V}_e(\theta, \theta'|\mathcal{M}^d) = U_e(\sigma_e', \sigma_\ell|\theta, \beta_e), \tag{30}$$

$$\mathcal{V}_e(\theta, \theta|\mathcal{M}^d) = U_e(\sigma|\theta, \beta_e), \text{ and} \tag{31}$$

$$\mathcal{V}_\ell(\mathcal{M}^d) = U_e(\sigma|\beta_\ell). \tag{32}$$

To see this suffices, note that the fact that (σ, β) is an individually rational PBE and Equations (30)-(32) imply that $\mathcal{M}^d$ is BIC and IR. Since transfers are also equivalent, the result follows from Proposition 4.2.

We now construct such strategy σ_e'. Set $\sigma_e'(\cdot)(I_e) = \sigma_e(\theta')(I_e)$ for each $I_e \in \Theta \times (\mathcal{I}_e \backslash \mathcal{T}_e)$ and $\sigma_e'(\cdot)(T_e) = \sigma_e(\cdot)(T_e)$ for each $T_e \in \mathcal{T}_e$. That is, σ_e' is a strategy profile such that the expert mimics an expert with state θ' in the mechanism $\mathcal{M}$ but plays according to σ in G.

To show Equation (30), we first show some identities. Notice that for each $\theta \in \Theta$ and each $T_\ell \in \mathcal{T}_\ell[\mu_\ell^1]$,

$$\mu(\theta) \cdot \mathbb{P}(T_\ell|\theta, \sigma) = \mathbb{P}(\theta, T_\ell|\sigma) = \mu_\ell^1(\theta) \cdot \mathbb{P}(T_\ell|\sigma), \tag{33}$$

where the first equality follows from definition and the second from the fact that the layman has posterior $\mu_\ell^1(\theta)$ of state θ at T_ℓ. By adding over $T_\ell \in \mathcal{T}_\ell[\mu_\ell^1]$, it follows that

$$\mu(\theta) \cdot \mathbb{P}(Z[\mu_\ell^1]|\theta, \sigma) = \mu_\ell^1(\theta) \cdot \mathbb{P}(Z[\mu_\ell^1]|\sigma). \tag{34}$$

Thus, if $\mathbb{P}(Z[\mu_\ell^1]|\theta, \sigma) > 0$, then $\mu_\ell^1(\theta) > 0$ and $\mathbb{P}(Z[\mu_\ell^1]|\sigma) > 0$. Moreover,

$$\begin{aligned}
\sum_{T_\ell \in \mathcal{T}_\ell[\mu_\ell^1]} \sigma_\ell(T_\ell) \cdot \frac{\mathbb{P}(T_\ell|\theta, \sigma)}{\mathbb{P}(Z[\mu_\ell^1]|\theta, \sigma)} &= \sum_{T_\ell \in \mathcal{T}_\ell[\mu_\ell^1]} \sigma_\ell(T_\ell) \cdot \frac{\mu(\theta) \cdot \mu_\ell^1(\theta) \cdot \mathbb{P}(T_\ell|\sigma)}{\mu(\theta) \cdot \mu_\ell^1(\theta) \cdot \mathbb{P}(Z[\mu_\ell^1]|\sigma)} \\
&= \sum_{T_\ell \in \mathcal{T}_\ell[\mu_\ell^1]} \sigma_\ell(T_\ell) \cdot \frac{\mathbb{P}(T_\ell|\sigma)}{\mathbb{P}(Z[\mu_\ell^1]|\sigma)} \\
&= q(\mu_e^1),
\end{aligned}$$

where the first equality follows form Equations (34) and (33) and the second from Equation (29). So, $\sum_{T_\ell \in \mathcal{T}_\ell[\mu_\ell^1]} \sigma_\ell(T_\ell) \cdot \mathbb{P}(T_\ell|\theta, \sigma) = q(\mu_e^1) \cdot \mathbb{P}(Z[\mu_\ell^1]|\sigma)$. This implies

$$\sum_{T_\ell \in \mathcal{T}_\ell[\mu_\ell^1]} \left(\int_{A_\ell} \pi_e(\theta, a_\ell) \; d\sigma_\ell(T_\ell) \right) \cdot \mathbb{P}(T_\ell|\theta, \sigma) = \left(\int_{A_\ell} \pi_e(\theta, a_\ell) \; dq(\mu_\ell^1) \right) \cdot \mathbb{P}(Z[\mu_\ell^1]|\theta, \sigma). \tag{35}$$

Notice, by construction σ' mimics θ in the mechanism. Hence, for each $T_\ell \in \mathcal{T}_\ell$ and each $\theta \in \Theta$,

$\mathbb{P}(T_\ell|\theta,(\sigma'_e,\sigma_\ell)) = \mathbb{P}(T_\ell|\theta',\sigma)$. Hence,

$$\begin{aligned}
U_e(\sigma'_e,\sigma_\ell|\theta) &= \sum_{\mu_\ell^1\in M_\ell^1}\sum_{T_\ell\in\mathcal{T}_\ell[\mu_\ell^1]}\left(\int_{A_\ell}\pi_e(\theta,a_\ell)\ d\sigma_\ell(T_\ell)\right)\cdot\mathbb{P}(T_\ell|\theta',\sigma)+\sum_{T_e\in\mathcal{T}_e}\gamma_e(T_e)\cdot\mathbb{P}(T_e|\theta',\sigma)\\
&= \sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_e(\theta,a_\ell)\ dq(\mu_\ell^1)\right)\cdot\mathbb{P}(Z[\mu_\ell^1]|\theta',\sigma)+\mathbb{E}Y_e(\theta'|\mathcal{M},\sigma)\\
&= \sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_\ell(\theta,a_e)\ q(\mu_\ell^1)\right)\cdot\operatorname{marg}_{M_\ell}m(\theta')(M_\ell^1[\mu_\ell^1])+\mathbb{E}Y_e(\theta'|\mathcal{M}^d,\rho^*)\\
&= \mathcal{V}(\theta,\theta'|\mathcal{M}^d),
\end{aligned}$$

where the second equality follows from Equation (35), the third by construction of $\mathcal{M}^d$, and the fourth from Lemma B.15. Hence, Equation (30) is satisfied and an analogous argument shows Equation (31).

To show, (32), notice that

$$\begin{aligned}
U_\ell(\sigma) &= \sum_{\theta\in\Theta}\sum_{\mu_\ell^1\in M_\ell^1}\sum_{T_\ell\in\mathcal{T}_\ell[\mu_\ell^1]}\left(\int_{A_\ell}\pi_\ell(\theta,a_\ell)\ d\sigma_\ell(T_\ell)\right)\cdot\mathbb{P}(\{\theta\}\times T_\ell|\sigma)+\sum_{T_\ell\in\mathcal{T}_\ell}\gamma_\ell(T_\ell)\cdot\mathbb{P}(T_\ell|\sigma)\\
&= \sum_{\theta\in\Theta}\sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_\ell(\theta,a_\ell)\ dq(\mu_\ell^1)\right)\cdot\mathbb{P}(\{\theta\}\times Z[\mu_\ell^1]|\sigma)+\mathbb{E}Y_\ell(\mathcal{M},\sigma)\\
&= \sum_{\theta\in\Theta}\sum_{\mu_\ell^1\in M_\ell^1}\left(\int_{A_\ell}\pi_\ell(\theta,a_e)\ q(\mu_\ell^1)\right)\cdot\operatorname{marg}_{\Theta\times M_\ell}\phi(\{\theta\}\times M_\ell^1[\mu_\ell^1])+\mathbb{E}Y_\ell(\mathcal{M}^d,\rho^*)\\
&= \mathcal{V}_\ell(\mathcal{M}^d),
\end{aligned}$$

where the third equality follows from construction of $\mathcal{M}^d$, and the last from Lemma B.15. ☐

Proposition B.3. *The game G of the game from Example 8.1 has no reduced form representation.*

Proof. First notice that there is a supergame $(\mathcal{M},G)$ and a PBE thereof that provides expected payoffs of $(5,5)$. To see this, define $\mathcal{M}$ as follows: Chance selects an element of $\{(\overline{a}_e,\underline{a}_\ell),(\overline{a}_e,\overline{a}_\ell),(\underline{a}_e,\overline{a}_\ell)\}$ with equal probability. Once chance selects (a_e,a_ℓ) each agent i privately observes their component a_i. There are no transfers. Let (σ,β) be a consistent profile such that follow the mechanism's recommendation. Notice that no agent has incentives to deviate and thus, (σ,β) is a PBE of $(\mathcal{M},G)$ and provides expected payoffs of 5 to each agent.

Since $\Theta=\{\theta\}$ is a singleton, the belief structure $H_e\times H_\ell=\{h_e^\theta\}\times\{h_\ell^\theta\}$ is also a singleton. Thus, to show that G has not reduced form representation, it suffices to show that (u_e,u_ℓ) with $(u_e(\theta,h_e^\theta),u_\ell(h_\ell^\theta))=(5,5)$ is not a reduced form. We show this by contradiction.

Assume that (u_e,u_ℓ) is a reduced form with $(u_e(\theta,h_e^\theta),u_\ell(h_\ell^\theta))=(5,5)$. Fix a mechanism $\mathcal{M}=(\cdot,\mathcal{T}_e,\mathcal{T}_\ell)$ with $\mathcal{T}_e=\{T_e\}$ and $\mathcal{T}_\ell=\{T_\ell\}$ both singletons, and $\beta\in\operatorname{Cons}(\mathcal{M})$. Consider the

Bayesian game $BG(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$. Since (u_e, u_ℓ) is a reduced form, there is a Bayesian equilibrium $\widetilde{\sigma}$ of $\mathrm{BG}(\mathcal{T}_e, \mathcal{T}_\ell, \beta)$ such that:

(i) $\Pi_e(\widetilde{\sigma} \mid \theta, T_e, \beta_e) = u_e(\theta, h_e^\theta) = 5$.

(ii) $\Pi_\ell(\widetilde{\sigma} \mid T_\ell, \beta_\ell) = u_\ell(h_\ell^\theta) = 5$.

Since $\mathcal{T}_e$ and $\mathcal{T}_\ell$ are both singletons, this implies that G has a Bayesian equilibrium with payoffs $(5, 5)$, leading to a contradiction. □

Lemma B.16. *The game G from Example 8.2 has a reduced form representation* $\mathrm{RM} = \{(u_e, u_\ell)\}$ *such that u_e is not supermodular in common degenerate beliefs and is not submodular with respect to any acute and essential statistic f. Moreover, G is not perfectly revealing or concealing.*

Proof. Analogously to Proposition 7.2, there is a reduced form representation $\mathrm{RF} = \{(u_e, u_\ell)\}$ such that

$$u_e(\theta, h_e) = -\int\limits_{h_\ell \in H_\ell} \left(\mathrm{b}(\theta) - \mathbb{E}\theta_\ell^1(h_\ell)\right)^2 d\eta_e(h_e), \qquad \text{and}$$

$$u_\ell(h_\ell) = -\sum_{\theta\in\Theta} \left(\theta - \mathbb{E}\theta_\ell^1(h_\ell)\right)^2 \operatorname{marg}_\Theta \eta_\ell(h_\ell)(\theta).$$

Since $b(\cdot)$ is decreasing in some parts, the function u_e does not satisfies cyclical monotonicity on degenerate beliefs (and hence is not supermodular on degenerate beliefs). To see this, write $g(\theta, \theta') = u_e(\theta, h_e^{\theta'}) = -(\mathrm{b}(\theta) - \theta')^2$ and consider the cycle $\theta = (2, 3, 2)$. Then,

$$\mathcal{L}(g, \theta) = g(2, 2) - g(2, 3) + g(3, 3) - g(2, 3) = -1 + 0 - 1 + 0 < 0.$$

Hence, u_e does not satisfies cyclical monotonicity on degenerate beliefs. Moreover, this implies that G is not perfectly revealing. (See Theorem 8.1.)

To show that u_e does not satisfy the supermodularity condition, it suffices to show that G is not concealing. (See Theorem 5.2.). Thus, it suffices to show that there are mechanisms and equilibria that affect the action of the layman, and therefore, the agent's payoffs. To see this, consider the mechanism $\mathcal{M}$ where the expert selects on of two (public) cheap talk messages $\underline{m}$ and $\overline{m}$. Consider the strategy profile σ of the supergame $(\mathcal{M}, G)$ where the expert reveals if $\theta = 1$ or $\theta \in \{2, 3\}$ and the layman optimally reacts. That is, $\sigma = (\sigma_e, \sigma_\ell)$ is the (pure) strategy profile such that $\sigma_e(1) = \underline{m}$, $\sigma_e(2) = \sigma_e(3) = \overline{m}$, $\sigma_\ell(\underline{m}) = 1$, and $\sigma_\ell(\overline{m}) = 2.5$. Let β be beliefs with σ. Notice that no agent has incentives to deviate and therefore (σ, β) is a PBE. □

***Proof of Theorem** 8.1.* *If.* Fix $(u_e, u_\ell) \in \mathrm{RF}$ so that $g(\theta, \theta') = u_e(\theta, h_e^{\theta'})$ satisfies cyclical monotonicity. Then there is $z : \Theta \to \mathbb{R}$ such that $g(\theta, \theta) + z(\theta) \geq g(\theta, \theta') + z(\theta')$. (See Theorem 4.2.1 in Vohra [2011].) Thus, the construction of $\mathcal{M}^d$ from Theorem 5.1 can be applied to this environment and, as a consequence, G is perfectly revealing.

Only if. Assume G is perfectly revealing. Then, there is a mechanism $\mathcal{M}$ and a PBE (σ, β) of the supergame $(\mathcal{M}, G)$ where the layman learns the state. Since RF is a reduced form representation of G, there is a reduced-form $(u_e, u_\ell) \in \text{RF}$ and a credible, BIC direct mechanism $\mathcal{M}^d$ for (u_e, u_ℓ) with a set of message profiles $M = \text{CDB}_e \times \text{CDB}_\ell$. (See Proposition 4.2) write $g(\theta, \theta') = u_e(\theta, h_e^{\theta'})$, $z(\theta) = \sum_{y_e \in Y_e} y_e \text{marg}_{Y_e} m(\theta)(y_e)$ and note that $V(\theta, \theta'|\mathcal{M}^d) = g(\theta, \theta') + z(\theta')$. Thus, since $\mathcal{M}^d$ is BIC, it follows that $g(\theta, \theta) + z(\theta) \geq g(\theta, \theta') + z(\theta')$. Hence, Theorem 4.2.1 in Vohra [2011] implies that g satisfies cyclical monotonicity. □

To show Proposition 8.1, we introduce some notation. Fix the set of states $\Theta = \{0, 1\}$ and a direct mechanism $\mathcal{M}^d$. For each $\theta \in \Theta$ write $\nu_\theta \in \Delta([0,1])$ for the distribution of researcher's posteriors beliefs that the subject is of type 1 contingent on a report θ. That is, for each measurable $B \subseteq [0,1]$, $\nu_\theta(B) = \text{marg}_{M_\ell} m(\theta)(p^{-1}(B))$. Notice, since p is measurable, ν_θ is well defined.

Lemma B.17. *Fix a credible direct mechanism $\mathcal{M}^d$ with associated ex-ante measure ϕ. Then*

(i) $\text{marg}_{\Theta \times M_\ell} \phi(1, h_\ell) = p(h_\ell) \cdot \text{marg}_{M_\ell} \phi(h_\ell)$.

(ii) $\text{marg}_{\Theta \times M_\ell} \phi(1, h_\ell) = (1 - p(h_\ell)) \cdot \text{marg}_{M_\ell} \phi(h_\ell)$.

Proof. We show (i). (Showing (ii) follows from an analogous argument.)

$$\begin{aligned}
\text{marg}_{\Theta \times M_\ell} \phi(\theta, h_\ell) &= \sum_{h_e \in M_e} \sum_{y_\ell \in Y_\ell} \text{marg}_{\Theta \times M \times Y_\ell} \phi(\theta, h_e, h_\ell, y_\ell) \\
&= \sum_{h_e \in M_e} \sum_{y_\ell \in Y_\ell} \text{marg}_{\Theta \times M \times Y_\ell} \phi(\theta, h_e, h_\ell, y_\ell) \\
&= \sum_{h_e \in M_e} \sum_{y_\ell \in Y_\ell} p(h_\ell) \cdot \text{marg}_{\Theta \times M \times Y_\ell} \phi(h_e, h_\ell, y_\ell) \\
&= p(h_\ell) \cdot \text{marg}_{M_\ell} \phi(h_\ell),
\end{aligned}$$

where the third equality follows from credibility. □

Lemma B.18. *Fix a credible direct mechanism $\mathcal{M}^d$. Then, the measure ν_1 first-order stochastic dominates ν_0. That is, for each $k \in [0,1]$, $\nu_1[0,k] \leq \nu_0([0,k])$.*

Proof. Fix $k \in [0,1]$. Assume $k \leq \mu(1)$. (The case $k > \mu(1)$ is analogous.) Write $\underline{M_\ell} = \{h_\ell \in M_\ell : p(h_\ell) \in [0,k]\}$. Notice that

$$\sum_{h_\ell \in \underline{M_\ell}} \text{marg}_{\Theta \times M_\ell} \phi(\theta, h_\ell) = \sum_{h_\ell \in \underline{M_\ell}} \mu(\theta) \cdot \text{marg}_{M_\ell} m(\theta)(h_\ell) = \mu(\theta) \cdot \nu_\theta([0,k]),$$

Thus, to show $\nu_1([0,k]) \leq \nu_0([0,k])$, it suffices to show that

$$\mu(0) \cdot \sum_{h_\ell \in \underline{M_\ell}} \text{marg}_{\Theta \times M_\ell} \phi(1, h_\ell) \leq \mu(1) \cdot \sum_{h_\ell \in \underline{M_\ell}} \text{marg}_{\Theta \times M_\ell} \phi(0, h_\ell). \tag{36}$$

Notice, $k \leq \mu(1)$ implies

$$\sum_{h_\ell \in \underline{M_\ell}} p(h_\ell) \cdot \text{marg}_{M_\ell} \phi(h_\ell) \leq \sum_{h_\ell \in \underline{M_\ell}} \mu(1) \cdot \text{marg}_{M_\ell} \phi(h_\ell).$$

Since $\mu(0) = 1 - \mu(1)$, this holds if and only if

$$\sum_{h_\ell \in \underline{M_\ell}} \mu(0) \cdot p(h_\ell) \cdot \text{marg}_{M_\ell} \phi(h_\ell) \leq \sum_{h_\ell \in \underline{M_\ell}} \mu(1) \cdot (1 - p(h_\ell)) \cdot \text{marg}_{M_\ell} \phi(h_\ell).$$

Hence, by Lemma B.17,

$$\sum_{h_\ell \in \underline{M_\ell}} \mu(0) \cdot \text{marg}_{\Theta \times M_\ell} \phi(1, h_\ell) \leq \sum_{h_\ell \in \underline{M_\ell}} \mu(1) \cdot \text{marg}_{\Theta \times M_\ell} \phi(0, h_\ell).$$

So, Equation (36) holds. □

***Proof of Proposition** 8.1.* First we show (i). For simplicity we assume that the outside options of the agents are both zero. Notice that $\mathbb{E}u_e$ is weakly supermodular at common degenerate beliefs if and only if g satisfies weakly increasing differences on $\Theta \times \{0, 1\}$. Hence, the construction of Theorem 5.1 holds. There is a credible, BIC, IR direct mechanism where the layman learns the state under the honest equilibrium.

Now we show (ii). By the revelation principle in Rivera Mora [2021b]. It suffices to analyze only credible, BIC, and IR mechanisms. Fix a credible, BIC and IR direct mechanism $\mathcal{M}^d$ and let $\nu_\theta \in \Delta([0, 1])$ be such that $\nu_\theta([0, k]) = m(\theta)(\{h_\ell \in H_\ell : p(h_\ell) \leq k\})$. Notice that BIC implies $\mathcal{V}(1, 1|\mathcal{M}^d) - \mathcal{V}(1, 0|\mathcal{M}^d) \geq 0 \geq \mathcal{V}(0, 1|\mathcal{M}^d) - \mathcal{V}(0, 0|\mathcal{M}^d)$. Thus,

$$\begin{aligned}\int_0^1 (g(1, p) - g(0, p)) \, d\nu_1 &= \mathcal{V}(1, 1|\mathcal{M}^d) - \mathcal{V}(0, 1|\mathcal{M}^d) \\ &\geq \mathcal{V}(1, 0|\mathcal{M}^d) - \mathcal{V}(0, 0|\mathcal{M}^d) \\ &= \int_0^1 (g(1, p) - g(0, p)) \, d\nu_0 \end{aligned} \tag{37}$$

Nos, suppose that $\nu_0 \neq \nu_1$. By Lemma B.18, ν_1 first-order stochastically dominates ν_0. Moreover, by strict decreasing differences, $\hat{g}(p) := g(1, p) - g(0, p)$ is strictly decreasing in p. Thus $\nu_0 < \nu_1$ implies $\int_0^1 \hat{g}(p) \, d\nu_1 < \int_0^1 \hat{g}(p) \, d\nu_0$ (See Border [2001]). This contradicts Equation (37). Therefore, $\nu_0 = \nu_1$. Write $f : H \to \mathbb{R}$ given by $f(h_e, h_\ell) = p(h_\ell)$. Notice that f is acute. (See Lemma B.5.) Let $\mathbf{F}$ be the random variable associated to f. Since $\nu_0 = \nu_1$, the state θ does not impact the layman's posterior. So, $\text{Cov}_\phi[\mathbf{F}, \boldsymbol{\Theta}] = 0$ and thus $\mathbf{F}$ is equal to the prior value $p(\tilde{h}_\ell)$. So, layman's posterior is equal to the prior. □